THE DEVELOPMENT OF TECHNIQUES FOR ORGANIC PROFILING OF GEOFORENSIC TRACE EVIDENCE

Georgia McCulloch

Thesis submitted in partial fulfilment of the requirements for the degree of Doctor of Philosophy

UNIVERSITY COLLEGE LONDON (UCL)

Declaration

I, Georgia M^{c}Culloch, hereby declare that this dissertation is my own original work and that all source material used has been clearly identified and acknowledged. No part of this dissertation contains material previously submitted to the examiners of this or any other University, or any material previously submitted for any other examination.

Signed

Date

Abstract

There is a need to expand the range of analytical techniques which can be used to analyse the organic components of geoforensic trace evidence in order to allow the discrimination of more forensically relevant soils, situated at smaller distances from one another. Existing analytical techniques used in forensic geoscience are based on the analysis of the inorganic fraction of soil, which is often consistent across several kilometers, therefore a complementary analytical technique is required in order to maximise the evidential value of soils and sediments. HPLC has been proposed here as a suitable method, and during method development the most appropriate methods for handling and analysing soil evidence by HPLC were selected, producing a method that is significantly more sensitive, less costly, complicated and time consuming than the methods proposed by previous authors. The feasibility of reliably excluding soil samples from locations situated approximately 250m apart using the new HPLC method was demonstrated, and 100% accuracy ($p=0.000$) was obtained by performing canonical discriminant function analysis on the resulting HPLC data. The highly complex chromatograms obtained present challenges during data analysis which were resolved by selecting two subsets of markers, each containing a reduced number of peaks than the initial data set. This experiment was repeated at three UK sites, over 12-18 months, and at one site in the USA, and HPLC was shown to offer extremely high accuracy rates in discriminating samples at all four sites and therefore to be robust to changes in underlying geology, and to provide highly accurate discrimination at all time points, and to be robust to delays between sample collection and analysis. Collaborative experiments were undertaken, in which HPLC samples were re-analysed using quartz grain surface texture analysis, and wax marker profiling by GC, and HPLC was shown to offer better discrimination at this close-proximity spatial scale, in addition to offering potential benefits in regulated and commercialised forensic science laboratories due to its simplicity and efficiency.

Acknowledgements

Tremendous thanks are due, first and foremost, to my supervisor, Dr Ruth Morgan, for her unwavering optimism, tenacity, and patience in helping me to complete this project, in particular, her guidance and moral support with writing the thesis, and her invaluable technical expertise in quartz grain surface analysis. Thank-you for keeping me sane, sticking with me and being that reassuring voice on the end of the phone during my "wobbly moments".

In addition, I would like to acknowledge the funding for the project provided by the Engineering and Physical Sciences Research Council through the Security Science Doctoral Research Training Centre (UCL SECReT) based at University College London (EP/G037264/1) and those contributions made by Dr Adrian Jones and Professor Simon Gibbons as secondary supervisors.

Professor Lorna Dawson of the James Hutton Institute is owed particular gratitude for her advice and for facilitating access to HPLC and GC equipment and laboratory facilities, without which this research would have been impossible. I would like to thank Dr Peter Bull and the Oxford University Centre for the Environment for offering the expert consultation and laboratory facilities used during the quartz grain surface texture analysis experiments. Likewise, I am extremely grateful to Dr Mark Brewer of BioSS for performing the statistical analysis in "R" and advice on improvements to the data analysis. Mrs Jasmine Ross, Dr Robert Mayes, and Mrs Sheila Reid of the James Hutton Institute were extremely kind and helpful and I am thankful for their advice and assistance in conducting the wax marker analyses.

I am extremely grateful to my former colleague Katy Basford for generously donating the columns and consumables, which allowed the scope of the project to be expanded significantly. My wonderful partner, Gareth Foley was critical to the validity and potential impact of this research in his role as second analyst, and without his emotional and financial support during the final stages, the project could not have been completed. To my Granny Joy and all the infuriatingly clever McCullochs, thank-you for giving me the drive to keep going.

Table of Contents

List of Figures

List of Tables

Glossary

AAS	Atomic Absorption Spectroscopy
AES	Atomic Emission Spectroscopy
°C	Degrees Centigrade
CDFA	Canonical Discriminant Function Analysis
C18	Octadecyl Silanol
cGCLP	Current Good Clinical Laboratory Practice
DAD	Diode Array Detector
cm	Centimetres
ESI	Electrospray Ionisation

df	Degrees of Freedom
FTIR	Fourier Transform Infra-red
g	Grams
GC	Gas Chromatography
HPLC	High Performance Liquid Chromatography
ICP	Inductively Coupled Plasma
km	Kilometre
LC	Liquid Chromatography
LDPE	Low Density Polyethylene
MeCN	Acetonitrile
MeOH	Methanol
mg	Milligram
MI	Michigan
min	Minutes
ml	Millilitres
mm	Millimetres
MS	Mass Spectrometry
N/A	Not Applicable
N/D	Not Detected
nm	Nanometres
ODS	Octadecyl Silanol
pH	Measure of the acidity of a sample
PTFE	Polytetrafluoroethylene
RPM	Revolutions Per Minute

RT	Retention Time
S	Seconds
SEM	Scanning Electron Microscope
Sig	Significance
UPLC	Ultra Performance Liquid Chromatography
UV	Ultraviolet
V	Volts
Vis	Visible
XRD	X-Ray Diffraction
XRF	X-Ray Fluorescence
λ_{max}	Lambda max
µl	Microliters
µm	Micrometers
µV	Micro-Volts

1 Introduction

1.1 Introduction

Forensic Geoscience applies the techniques developed to study earth materials as it pertains to the courts. Earth materials include rocks, sediments and soils and their chemical, biological and anthropogenic components. While forensic geoscience must, of course, pertain to the law, the subject is not limited in scope to assisting police work, indeed it has applications in any legal context where earth materials may be able to help investigators, judges or jurors establish "what happened, where and when it occurred and how and why it took place"(Ruffell and McKinley, 2008:p1) (1). Since earth materials are highly transferable, persistent and present at a wide variety of crime scenes, geo-forensics can be used in a multitude of scenarios to aid crime reconstruction, corroborate witness statements or verify suspect alibis (2).

Trace geo-forensic evidence recovered from a suspect, victim or crime scene can be interpreted in line with Locard's exchange principle, which can be generalised as "whenever two objects come into contact, there is always a transfer of material" (3), in order to establish whether there is a potential link between items or locations of forensic interest. Trace geo-forensic evidence can be used in a predictive manner in order to guide investigators to the location of its source, through the use of databases and maps in concert with expert local knowledge (4,5,6). Alternatively, earth materials from known and questioned samples can be compared in order to exclude or include the possibility of a common source.

Due to the complexity of the structure and composition of geo-forensic materials, which provides multiple criteria for comparison, there are a number of different ways in which geo-forensic trace evidence can be described and classified (7,8). Typically, the techniques used to do this have focused on the physical or chemical characteristics of the mineral fraction of soil (9,10,11), and are well established methods within the earth sciences for the purpose of studying the processes and events involved in the formation of the earth (10). In the past, geo-forensic analyses have been approached in much the same way as traditional geological analyses, arguably without due consideration of the practical and philosophical differences between the forensic and earth sciences (12,13,14,15). It could, therefore, be considered useful to develop analytical methods with the specific requirements of forensic casework borne in mind (16).

Geo-forensic samples can occur in a variety of forms and a number of matrices, and it cannot be guaranteed that these will be compatible with the sample handling and analysis methods

used in traditional geology. It is recommended that forensic evidence is tested with techniques that analyse independent characteristics of the sample in order to provide maximum weight to the conclusions drawn (13,14,15). In addition to the weathered rock fragments, which currently attract the greatest scrutiny in geo-forensic analyses, soil is comprised of biological material, liquid and gaseous components (17), therefore there is a need for more techniques capable for the analysis of soil components other than the inorganic minerals (9,18).

Soil forms layers, or horizons, which have different physical, chemical and biological properties depending on the type of soil and environmental conditions, as shown in figure 1.1.

Figure 1.1 Diagram showing the O, A, B and C soil horizons (19)

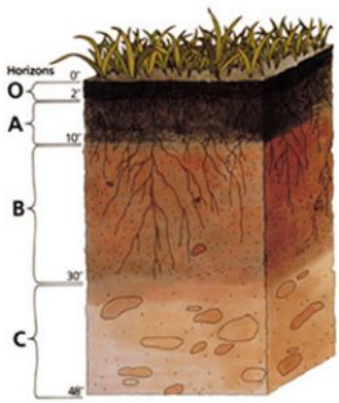


Figure 1.1 shows the uppermost layers of soils, the O and A horizons, are richer in organic matter (20,21), which is comprised of living organisms, their intact remains and the organic compounds (organic compounds, with very few exceptions, are those which contain carbon-hydrogen bonds and need not be biogenic) produced by their decomposition plus any synthetic organic compounds added to the soil

There are a number of techniques designed to analyse organic compounds, are primarily concerned with the separation, identification and quantification of organic compounds and techniques exist which are capable of handling solid, liquid and gaseous samples. Appropriate organic analytical techniques are well established within forensic science since these compounds are found in fibres (22), explosives (23,24), accelerants (25), alcohol (26,27), drugs and poisons (27,28,29)

High performance liquid chromatography (HPLC) is a widely used organic analytical technique, which has been reported to be potentially useful in the analysis of forensic soil samples (18,30,31,32). The purpose of this study was to redevelop the sample collection, preparation, analysis and interpretative approaches used in order to assess the feasibility of the technique for use in a more forensically relevant context. The developed method was intended to be appropriate for comparing trace soil samples for the purposes of excluding crime scene, alibi site and unknown samples in criminal cases. These contextual details are important, since they affect the considerations required for correct interpretation of the evidence. The question this research addresses is:

'To what extent is it possible to distinguish groups of trace soil samples obtained from locations that are located in close proximity to one another using HPLC analysis?'.

In order to develop the previous work and address the gaps in the forensic geoscience literature identified and discussed in Chapter 2, this study aims to develop an HPLC method for the discrimination of close proximity, trace soil and sediment samples for use as forensic evidence and intelligence.

1.2 Thesis Outline

This thesis is comprised of nine chapters, as summarised below.

1.2.1 Chapter 2

The literature review in Chapter 2 outlines the conceptual foundations for this research of the forensic process from the crime scene, through the analysis of evidence, and the interpretation of analytical results, to the final presentation of evidence in court. A background to the history of the provision of forensic science services in the UK is provided followed by a summary of the underlying scientific theory of the soil forming processes that are used in forensic geoscience. The techniques used previously in forensic geoscience research and casework are also presented, alongside a discussion of the limitations of these techniques and the areas in which research into the development of new analytical methods

could potentially add value. In addition, the philosophical principles that are crucial to undertaking sound scientific research within the forensic scienc domain and improving the reliability of forensic evidence are examined. The techniques commonly used to identify and quantify organic compounds are discussed and the chemical mechanisms that result in the chromatographic purification and separation of sample mixtures during HPLC analyses are presented. The operation of an HPLC instrument is also explained, and examples of the application of HPLC in forensic science research generally, as well as forensic geoscience specifically, are provided, with a discussion of the areas in which these studies lack validity and could be improved upon in the current research. Lastly, the aims of the thesis are articulated.

1.2.2 Chapter 3

In chapter 3, a series of method development experiments are presented in which the sample preparation and handling procedure and HPLC parameters are optimised. The aim of the research presented in this chapter was to improve upon the practicality of the methodology proposed by previous authors, and to develop a user-friendly method, appropriate for use in a commercial laboratory, by simplifying the sample preparation method, reducing the time required to complete the analysis, and reducing the quantity of soil required relative to previous studies. A variety of analytical solvents and mobile phase gradients were trialled, in addition to various different types of HPLC column, resulting in a reduction in the run time in relation to previous research. The sample preparation method is simplified, reducing the sample preparation time and the cost of the technique, and the amount of soil required for each analysis is reduced to 250mg. In addition, the stability of the prepared samples was investigated and the effect of different types of packaging and storage conditions monitored. Two new, practical methods for sample handling, preparation and HPLC analysis are presented, which are used to generate the results of chapters 4-7, and method development for the initial analytical method has been presented by McCulloch et. al. (33)

1.2.3 Chapter 4

The aim of chapter 4 was to determine the ability of one of the new methods, developed in chapter 3, to discriminate samples from close-proximity locations in order to demonstrate the feasibility of developing HPLC for use as a tool for forensic geoscience. Samples were collected from close-proximity locations within a discrete sample site at Brockwell Park in London and analysed using one of the new methods developed in chapter 3, 100% accuracy in grouping samples according to provenance was achieved in this study indicating that the technique is well suited to discriminating close-proximity trace soil samples on an exclusionary basis. This feasibility study has also been presented by McCulloch et. al. (33)

1.2.4 Chapter 5

The aim of chapter 5 was to reduce the time taken to complete each analysis. This chapter describes the various data analysis strategies used in the thesis and the steps taken to improve the speed with which results can be delivered using the new HPLC method. The data produced from samples collected from four different sites was analysed by visual comparison and statistical analysis and two sets of marker peaks are identified that are capable of producing consistently high rates of accuracy, reducing the time taken to complete the data analysis from several weeks to less than five minutes, significantly improving the practicality and cost of performing the analysis and, therefore improving the suitability of the method for use in a commercial laboratory.

Some of the work to identify one of the peak marker sets used the R "Subselect" package, and this step was performed at the James Hutton Institute by Dr Mark Brewer of Biomathematics and Statistics Scotland (BioSS). Aspects of this work have been submitted for publication in Forensic Science International (34)

1.2.5 Chapter 6

The aim of chapter 6 was to apply the technique to samples from locations representing a range of underlying geologies, and to monitor temporal variations in soil chromatography in order to verify the ability of the technique to provide accurate results at different times of year and in different locations, and identify the effects of delayed sample collection on the ability to reliably discriminate samples. Samples from four sites, from the UK and USA, are collected across an 18 month timescale and analysed in chapter 6 using the new methodology and marker sets, and high rates of accuracy are achieved in all of the discriminant analyses, and in addition, one of the peak sets identified in chapter 5 is shown to be robust to delays in sample collection, suggesting the suitability and practicality of the new HPLC method for general use. The results obtained for the four sample sites at the winter time point have been presented by McCulloch et. al. (34) (35)

1.2.6 Chapter 7

The aim of chapter 7 was to compare the discriminatory power of the method with existing geoforensic techniques in order to determine the ability of the new HPLC to contribute to the existing suite of tools for forensic geoscience. Samples were analysed using the new HPLC technique, and by GC to determine the wax marker profiles, which is the established technique for analysing the organic fraction of soil in forensic investigations, and with Quartz Grain Surface Texture Analysis (QGSTA), which is one of the established techniques for analysing the inorganic fraction of forensic soils and sediments. The new HPLC technique was found to be

substantially quicker than both existing techniques, and far less complicated than the GC technique, and shown to provide better discrimination of close proximity samples than both the established techniques. The results of this chapter suggest that combined application of the new technique alongside existing techniques can be beneficial in terms of the cost, speed, practicality and overall accuracy of the analyses, in addition to strengthening the evidential value of forensic soil traces when used with independent techniques such as QGSTA. The comparison of HPLC and QGSTA has also been presented by McCulloch et. al. (33) while the comparison of the results obtained by GC and HPLC have also been submitted for publication (35) and Dr Peter Bull and Dr Ruth Morgan both assisted with sample preparation and the assessment of the grain type characteristics for the QGSTA.

1.2.7 Chapter 8

Chapter 8 presents a summary of the key themes of the thesis and summarises the main findings of the research, and their implications on current forensic geoscience capabilities. In addition, potential avenues for future research are outlined.

1.2.8 Chapter 9

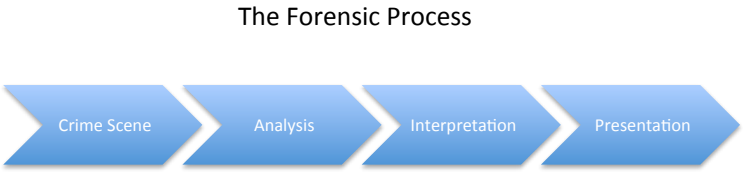
Chapter 9 presents the conclusions of the thesis, and an assessment of the extent to which the seven objectives, provided in chapter 2, have been achieved

2 Literature Review

2.1 The Forensic Process

Forensic Science is a broad field, encompassing all scientific endeavours to assist legal proceedings, both criminal and civil, and though the techniques used in forensic science that originate from a range of different disciplines. All forensic science follows the same process and lies within the same conceptual framework. The forensic process was traditionally considered to encompass those activities occurring at the crime scene, the analysis of evidence, interpretation of analytical data, and presentation of the results to assist the triers of fact in court, as shown in figure 2.1, and was based on five basic concepts: Transfer; Identification; Individualisation; Association; Reconstruction (36,37,38,39,40).

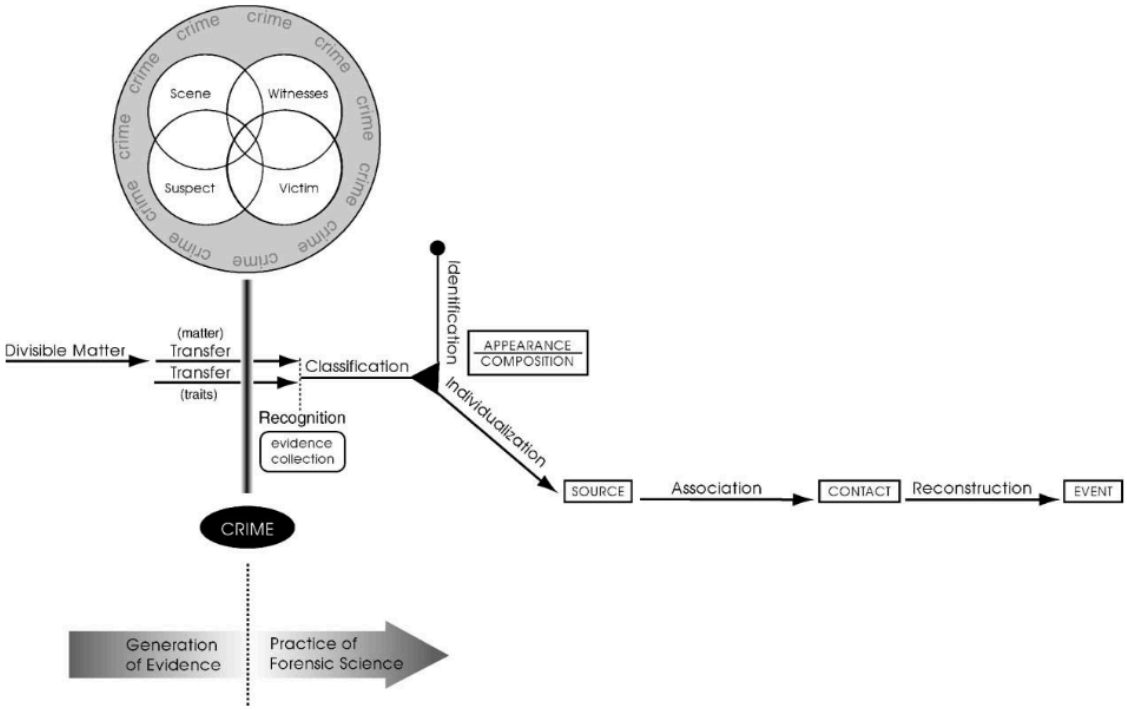
Figure 2.1 Process diagram representing the simplified forensic process.



More recently, however, this framework was expanded by Inman and Rudin (41), and reiterated by Morgan and Bull (13), to include the division of matter as a sixth central tenet of the forensic process, which is illustrated in figure 2.2. In this updated forensic paradigm, the division of matter must happen before the transfer of evidence can occur.

The schematic provided by Inman and Rudin (41), and shown in figure 2.2, depicts the different phases through which evidence passes, and the complexity of processes occurring on the journey from crime scene to court. As the division and transfer of matter are processes occurring continuously for all events, whether of forensic interest or not, the processes that are specifically relevant to forensic investigations can be simplified as crime scene, analysis, interpretation and presentation, which is summarised in figure 2.1. The experimental research in this thesis broadly pertains to the analysis stage of the forensic process, however since all of the stages impact on one another and are interrelated, all of the practical work undertaken for this research is carried out in the context of its position within the holistic process, thus the research has relevance to the crime scene, interpretation, or presentation stages of the process.

Figure 2.2 Schematic diagram representing the forensic process as presented by Inman and Rudin (41)



2.1.1 The Division, Transfer and Persistence of Matter

The first step, the division of matter occurs when a physical action, and the energy created by the action, cause a discrete object to break apart into smaller constituents. The act of division itself may, for some materials, under very specific circumstances create features that allow the unambiguous association of the constituent parts to the larger object, for instance a physical "match" between a fragment of broken glass and the remainder of the bottle it came from (41). In some cases, the products of the division of matter will retain some properties which are so complex or rare that they are unique to the original object allowing the components to be linked to their specific source, for instance the DNA profile extracted from a cell may be sufficiently complex to attribute the source to the individual who shed that cell. Depending on the strength of the forces involved, however, the division of matter may also result in the loss of some of the observable, characteristic properties of the original object, for instance the resulting constituents may be so small in size that the shape and structure of the original item are no longer observable, hindering elucidation of their specific origin, however, for instance, the chemical composition of the parts is likely to remain unchanged, particularly for manufactured items, allowing for the small constituent parts to be attributed to any object with the same composition, including the specific source object. The extent to which these properties are retained is directly dependant on the degree of homogeneity of the original object. After division, the physical environment and forces acting on the original source object

and the smaller constituents will be likely to be different, potentially causing changes to the structure and composition that vary for the source object and each of the constituent parts. Any differences arising as a result of the consequences of the division of the matter, may create differences between the analytical data generated from the original object, and the data from the constituents. The division of matter is therefore an important addition to the conceptual framework used in the analysis, interpretation and discussion of forensic evidence since it enables appropriate consideration of the relationship between the properties of the proposed source material and the smaller constituents produced from the division of matter recovered from a crime scene (41,13)

After the division of matter has occurred, material can then be transferred to other surfaces and objects. Locard's exchange principle that "every contact leaves a trace" states that "whenever two objects come into contact, there is always a transfer of material. The methods of detection may not be sensitive enough to demonstrate this, or the decay rate may be so rapid that all evidence of transfer has vanished after a given time, nonetheless the transfer has taken place" (42)(Murray and Tedrow, 1992:p7). It is argued that it is this principle that provides the underlying foundation for the forensic sciences, and it is this principle that undergirds all the inferences that are made about materials within crime reconstructions. The division and transfer of matter happens in all contact situations between objects and is not restricted to forensic events. Therefore it is essential to consider that the transfer of material to an object, which is later used as forensic evidence, may have occurred before, during or after a crime event and that any traces of material recovered are likely to be a mixture of the results of these pre-, syn- and post-forensic transfers (13) (14) (43). This has important implications for analytical techniques that generate complex profiles which are then used to individualise or determine the provenance of a trace, as these profiles may be the result of a pure, single source or the combination of the profiles of an unknown number of sources, mixed together in unknown ratios. Another important consideration is that once the initial primary transfer occurs as a result of contact between two objects, for instance the transfer of fibres from an item of clothing to an item of furniture and vice versa, and each of these objects may subsequently come into contact with new objects, for instance the original item of clothing may be stored next to another garment, and in doing so there may be a secondary transfer of fibres from the furniture to the garment with which it has had no direct contact. These indirect transfers have been demonstrated to continue to occur as tertiary transfers, quaternary transfers and so on (44) and this type of interaction must be considered during interpretation of such trace evidence. Transfer of material may be one-way or two-way in nature so, for example, physical contact between an offender and a victim during an assault

may result in transfer of fibres from the clothing of an offender to the skin of the victim, likewise cells from the skin of the victim may be transferred to the clothing of an offender, both of which are examples of a one-way transfer, whereas it may be that both processes occur simultaneously, in a two-way transfer (14). In addition to transfers resulting from contact, there may be transfer through air, particularly for lightweight or particulate matter, without any contact being made, for instance a hair may be transferred onto clothing via air currents and gravity. Lastly, when considering the evidential value of transferred material, it is important to consider the size, shape, weight and texture of the objects and the forces involved in the transfer and subsequent use of the objects as these factors affect the extent to which transferred material is likely to be retained during the initial contact as well as the rate at which it is lost from the surface after transfer.

In light of the numerous factors related to the transfer of evidence, that may confound attempts to identify any specific source, there is a need for experimental research into the transfer and persistence of a range of types of physical evidence on a variety of different surfaces, that is relevant to the specific circumstances encountered in forensic casework, in order to build up empirical data against which to assess the significance of forensic test results (16).

2.1.2 Detection and Sampling

The next step in the forensic process is the first point at which the mechanisms involved become a matter of importance for forensic inference and crime reconstruction, since the division and transfer of matter are not limited to forensic scenarios. Once a crime is detected, investigators begin to search for evidence at the crime scene, and in order for evidence to be collected, it must first be observed and recognised as potentially useful evidence and therefore it is essential that researchers continue to provide investigators with the tools with which to search for evidence at the crime scene, for instance improved light sources to aid visualisation of bodily fluids or finger marks (45). In addition, there is a need for research to investigate the mechanisms through which evidence is changed and transported during and following a crime, in order that important evidence can be located and recognised quickly, for instance it may not be possible to recover fragments from a bomb blast without knowing where such particles are deposited after an explosion (46), and likewise, it is not possible to collect a sample from an aged bloodstain without understanding the way the appearance of blood changes over time (47). Furthermore, in light of the multiple, complex ways in which matter may be indirectly transferred, it is necessary to research improved crime scene procedures and products, such as new types of packaging and evidence-based protocols on how evidence should be stored and

transported, which are specifically designed to reduce any risk to the integrity of the evidence, such as contamination or alteration of the properties of the evidence (48) (49) (50) (51) (52).

2.1.3 Analysis

Following collection of evidence at the crime scene, the evidence can then be analysed by forensic scientists, which typically involves measurement and assessment of the physical characteristics, such as the shape and structure, or the biological or chemical composition of the samples in order to determine their identity or to individualise their origin to a specific source. In forensic science, identification is the process of classifying evidence and broadly pertains to studies that attempt to answer the question "what is it?" (41)while individualisation searches for an answer to the more specific question of "which one is it?" (41) to place the item into a class of one.

In recent years, there has been pressure from the wider scientific community, governing bodies and the criminal justice system for more research to improve the analytical methods used by forensic scientists to analyse evidence, to ensure they are based on sound scientific principles (53) (54) (55) (56). The process of designing a new way to test a particular material is known as method development, which typically involves experiments to ensure that the proposed test material, or analyte, can be extracted from the expected sample medium, and that the analyte can be detected, identified, and quantified in a timely, cost effective and timely manner, and that the methods of storing, handling and preparing the samples for analysis do not change the composition of the analyte or affect the analytical result. Analytical scientists working in other disciplines, such as materials science or drug discovery, use a holistic, risk based approach to analysis, sometimes referred to as "quality by design" (57). This strategy acknowledges that it is philosophically impossible to assure the quality of a group of items through testing alone, without testing each and every item individually, and that it is therefore essential to ensure that all steps in the lifecycle of the test item are well understood and controlled in order to minimise any potential risk to the quality of that item or the accuracy of the analytical results (58).

For forensic analyses however, it is not possible for the analyst to completely control the lifecycle of the test item to eliminate potential risks to the accuracy of the test result, since accuracy is a measure of the agreement between the measured value and the true or accepted value, which can never be established for forensic samples. It is therefore important that method development accounts for this and that test methods are designed to be suitable for non-ideal samples that may be aged, degraded or contaminated, and that the results are robust to a variety of likely, but uncontrollable, sources of variability. One of the primary ways

in which forensic scientists can justify confidence in the results of their analyses is through validation of the methods used to obtain them. Validation is defined by the international standards organisation (the ISO) as "the confirmation by examination and provision of objective evidence that the particular requirements for a specific intended use are fulfilled" (ISO/IEC, 2005: Clause 5.4.5.1) (59)and by the UK forensic science regulator (the Regulator) as "The process of providing objective evidence that a method, process or device is fit for the specific purpose intended." (The Forensic Science Regulator, 2014:6) (60) There are a number of indicators of whether a method is fit for purpose that should be verified during method validation, that depend on the requirements of the end user and the purpose of the analysis, and the list of issues which must be considered when validating forensic sampling and analysis methods, detailed by the International Laboratory Accreditation Cooperation's guidelines on the forensic science process, contains the following: Intended purpose and limitations; sampling strategy; sample homogeneity; accuracy; precision; measurement uncertainty; matrix effects ; interference; limit of detection; limit of quantification; linearity range; stability of measured compounds; specificity and selectivity; repeatability; reproducibility; robustness (61).

The first step of method validation in the framework set out by the Regulator is to define the scope of the analysis, which involves outlining the specific circumstances in which the analysis is intended to be used, and for what purpose (60). It is therefore not possible to validate an analytical method without careful consideration of how the samples have first been transferred and collected and how the resulting data will be used in crime reconstruction and presented in court.

2.1.4 Interpretation and Presentation

Following testing in the forensic laboratory, the analytical data must be interpreted correctly in order to enable crime reconstruction and to ensure that the evidence is presented accurately in court. Evidence is analysed with the ultimate aim of making comparisons between the recovered evidence, the source of which is unknown or questioned, and another object of known source that is relevant to the forensic case, for instance a reference sample collected from a suspect or the crime scene, therefore the analytical techniques used must be able to discriminate between items of different provenance in order to exclude potential common sources and interpret the data correctly. Such comparisons may be made against reference samples collected specifically for a particular case, or through the use of a database of samples collected previously, and it is essential that comparisons are made on a like-with-like basis in order to correctly exclude potential common sources (12). In order that databases can be used effectively in forensic cases, the reference samples in the database must be

representative of the source material (62) and must have been subject to the same conditions and treatment as the evidential sample, which not only impacts on the way evidential samples should be collected, handled and analysed in order to maintain consistency with reference samples but also limits the use of databases to cases where evidential samples have been or can be collected and tested in the same manner as the reference samples. While it is possible to construct large databases of relevant reference samples for evidence such as DNA, which allows for probabilistic interpretation of the resulting data through comparison with population based statistics on the frequency of alleles, for many types of evidence, including geoforensic evidence, there has been insufficient existing empirical data generated using experimental designs that are pertinent to specific case circumstances to allow comparisons to be made with such a high degree of statistical significance, and therefore interpretation for these types of evidence are made with reference to a small case-specific database of samples, on an exclusionary basis (13) (14).

It is therefore important to conduct research into the ways in which the properties and characteristic features of evidential samples may have been altered from the source material during the various stages of the forensic process, for instance the ways in which material may have been lost through transfers to other surfaces, or degraded over exposure to environmental or storage conditions prior to analysis. Due to the number of factors that may influence the properties of an item of evidence after division from its source, there is a need for experimental studies that replicate forensic reality as closely as possible, in order to build up a body of empirical data to support the theoretical models of evidence dynamics required to make accurate comparisons (41).

The extent of research required depends upon whether the evidence is to be interpreted at source level or activity level, and the specific context of a particular case (63) (64) (65) (14) (13). Source level determinations aim only to identify or individualise the specific object or class of material from which the evidence originates, in order to assess the possibility of a common origin with other items of evidence and aid the reconstruction of events (63) (64) (65). Interpretation at the source level requires research into the uniformity of the content of the source material in order to generate empirically justified sampling plans and ensure reference samples are representative of the particular source material in order to establish the range and variability of shared characteristics amongst all samples from the same source and permit the correct exclusion of samples from different locations. In addition, the impact of forensically relevant packaging and storage conditions on the composition of the source material must be understood through testing to verify the stability of the measured properties of the evidence under the conditions typically encountered in casework, along with studies to

identify and control for any biases introduced by the sample preparation or test procedure that may affect the accuracy of exclusionary analysis, since samples of the same origin may appear different after storage under different environmental and therefore be falsely excluded. Activity level evidence is used to assess the likelihood of a particular action having occurred, given the discovery of particular piece of evidence, for example to infer from the recovery of glass particulates from an item of clothing whether or not the person wearing the garment smashed a bottle at the crime scene (63) (64) (65). In order to assess evidence at the activity level in this forensic context, in addition to source level chemical testing to ensure the glass fragments were of the same composition as the bottle, and DNA testing of the clothing to ascertain whether the suspect had worn the garment in question, it would also be necessary to run experimental studies to determine whether glass fragments of the observed size and shape would be transferred in the observed number upon smashing a bottle, and whether those particles would be expected to persist in the observed quantity on the specific type of fabric over the alleged interval of time between the commission of the crime and the apprehension of the suspect for the specific case.

The final stage in the forensic process addresses the issue that in order to fully interpret the significance of the recovered evidence and present the evidence accurately in court, it is also necessary to investigate the impact of any other case circumstances presented by the defence or prosecution counsel. For instance, a defence proposition of secondary transfer from the perpetrator to the accused, or the significance of not finding evidence in cases where there has been a delay in detecting the crime and recovering evidence or where the prosecution proposes that the suspect has attempted to conceal the crime, for instance through washing clothes to remove evidence. The suitability of forensic evidence for presentation in court is an essential consideration in forensic science research as the justice system in many jurisdictions has strict standards the admissibility of forensic testimony in the courtroom. Many of the recent reforms to the way forensic science is conducted, for instance the Daubert requirements in some states in the US (66), and the recommendations of the Law Commission's report on expert evidence (67), have been prompted by the concerns of the judiciary over the reliability of the evidence presented by forensic experts in court, particularly the requirements for a sound set of empirically backed scientific principles to justify the opinion of forensic experts, and the need for a research culture within the forensic science community. Expert testimony in court is opinion based, and as a result it is essential that these opinions are informed by empirical research and data that aids the reconstruction of events to the jury, and accurately reflects the significance of each piece of evidence in the specific case context. Forensic science research must therefore aim to produce evidence that is suitable for

presentation in court by ensuring there is a demonstrable link between the available evidence and the opinion of the expert (66) (67) (68). In order to do so it must be shown that the theories used by forensic experts presenting their opinion in court are generally accepted in the scientific community, for instance through experimental studies, peer review and discussion in forensic science journals, and that all the techniques and processes used to produce the data upon which the opinion has been formed have been validated for the specific purpose for which they are used in court, in line with the relevant regulatory requirements of any given jurisdiction (53) (60) (66) (68) (67).

In the US, under the Daubert standard (66) there are four statutory criteria which must be fulfilled in order that opinion evidence is presented correctly in court by forensic experts, and therefore must be considered in forensic science research to ensure that the resulting evidence is admissible in court (69): Firstly, that the expert's scientific, technical, or other specialised knowledge will help the trier of fact to understand the evidence or to determine a fact in issue; secondly, that the testimony is based on sufficient facts or data; thirdly, that the testimony is the product of reliable principles and methods; and lastly the expert has reliably applied the principles and methods to the facts of the case. In order to satisfy these requirements, it is therefore necessary to demonstrate that the evidence is relevant to the case and the underlying theories have been applied reliably through correct interpretation (53) (60) (67) (68), for instance source level evidence is not relevant if the identity of the source object or person is not disputed (64) (63) (65). In order to demonstrate scientific validity and justify that forensic evidence is appropriate for presentation in court, it is also necessary that the methods and theories used to collect and interpret evidence are the product of the scientific method, which requires that theories are falsifiable through empirical testing, that the significance of the evidence generated by a particular technique can be assessed in the context of a known error rate specific to that evidence type and test, that there are up-to-date standards in place to ensure and verify the correct use of the particular forensic test, and that the technique is generally accepted within the scientific community through evidence of peer review and publication (53) (66) (60) (67) (69). Although no formal standards exist for the admissibility of forensic evidence in the UK court system, there have been calls from the government, practitioners and the Law Commission to develop a "gate-keeping" test for the validity of scientific expert testimony, that are based on the Daubert principles but are more specifically relevant to the requirements of UK legal and forensic science stakeholders, such as the judiciary, advocates and barristers, the police, and forensic science service providers (53) (67) (68) (70) (71).

2.2 UK Forensic Science Provision

In the UK, forensic science provision varies between the three legislative regions (England and Wales, Scotland, and Northern Ireland) and the degree of autonomy from police forces funding model has changed over the years (54) (55) (56) (68). In Northern Ireland and Scotland, forensic science is provided by public organisations affiliated with the national police forces (72) (73). In Scotland, forensic science services were historically provided in-house at four laboratories funded by the six individual police forces, however between 2007 and 2012 the provision of forensic services was formally separated from the police and became the responsibility of the Scottish Police Services Authority, which in turn became the current Scottish Police Authority, which is a public organisation funded by the Scottish Government, that provides central services to the Police, but that is autonomous from the constabulary. In Northern Ireland, forensic services are provided by Forensic Science Northern Ireland, which is also an autonomous publicly funded organisation, and is an agency of the Northern Irish Executive's Department for Justice.

In England and Wales, however the history of forensic science provision is more complex as the organisations providing forensic science services, and their funding mechanisms, have changed over time (71) (70) (72) (73) (68) (54) (55) (56). Since 1991 the majority of services were provided by the Forensic Science Service (FSS), which operated as a publicly funded agency of the UK Government's Home Office, however in 2005 the FSS became a government owned, contractor operated organisation known as a Gov.-Co, on the basis of the recommendations of the McFarland Review (74) (71) (70) (72) (73) (75). Legislative changes to the system through which police forces purchased goods and services were introduced in 1996 and forensic services were put out to competitive tender in an effort to reduce costs through increased competition amongst forensic science service providers (75) (75). This led to increased involvement of privately owned organisations in the provision of forensic science services, and a return to police forces in-sourcing their forensic services and by 2010 40% of forensic science services provided in England and Wales were provided by organisations other than the FSS. After transitioning to a Gov.-Co, it has been argued that the FSS management failed to make the necessary systemic changes required to operate sustainably in a competitive commercial environment, and in December 2010 the government announced their intention to close the FSS due to reported operational losses of £2 million per week (72) (73). The FSS closed in March 2012 and since then, forensic science provision has been undertaken by private companies and in-house by the police themselves. Following the closure concerns were voiced by members of the public, scientists and legal professionals prompted a series of reviews into the future of forensic science in the UK (54) (55) (56), and a

new national approach to forensic science provision and governance was announced in March 2016, which indicated that certain services such as fingerprint analysis, which is currently provided by individual police forces, would be conducted by a new, centralised organisation (76).

The changing nature of forensic science provision in the UK has led to concerns over the future of forensic science research in the UK due to the loss of investment in R&D that was previously provided by the FSS, the loss of certain specialised forensic services which are deemed unprofitable, and the loss of knowledge and expertise within the industry due to redundancies of the scientists working in these less commonly used or more expensive specialist areas (71) (70) (72) (73). As police budgets have been cut so too has their ability to pay for forensic services, this in turn negatively impacts on the profitability of the commercial forensic science providers who now provide the majority of forensic services in England and Wales, resulting in a lack of funding and motivation for investment in R&D within the private sector (71) (70) (72) (73). The majority of forensic science research in the UK must now be provided by the university sector, however as none of the primary research councils have a specific remit for forensic science, resources are also limited in academic research (55). Academic research into forensic science must therefore aim to continue to undertake the fundamental research traditionally provided in the context of university based research, in addition to the commercially orientated research conducted by practitioners and service providers and to meet the challenges of bridging the gap between primary research and commercialisation through casework relevant validation studies (71) (70) (72) (73) (68) (54) (55) (56). Since forensic science provision is now primarily a business in the UK, and with decision making on the utilisation of forensic science services being dictated by affordability and value for money, in order that new research can make an impact on practitioners it is essential to consider not only the ability of a technique to provide highly accurate and precise results for example for the identification of a particular source material, but also the cost of implementing the technique in the existing landscape of forensic science service provision. In order to assess the cost of commercialising a novel technique or device, it is necessary to consider operational costs of running the analysis, for instance the reagents and consumables required, the cost of performing the many and complex experiments needed to validate the procedure for use in a specific laboratory, the cost of purchasing, installing and maintaining any equipment used, and the salary costs of the scientist preparing samples and interpret data. With these commercial factors in mind, it can be seen that the most complex, state-of-the-art, specialist techniques may not always offer the best return on investment for UK forensic science providers, many of whom are small to medium enterprises (SMEs), and there is a need to develop new, efficient,

potentially profitable techniques that capitalise on existing capabilities and expertise (48) (54) (55) (56) (68) (71) (70) (72) (73).

2.3 The Origin and Diversity of Soil

There are over 700 soil types in the National Soil Map of England and Wales (77) and approximately 600 types of soil on the Scottish Soils Knowledge and Information Base (78), while there are 50,000 types of soil in the USA, and each soil type has its own specific set of genetic factors (79). The process of soil formation, pedogenesis (80), which is illustrated in figure 2.3, is considered to be the result of a combination of the five "clorpt" factors acting in conjunction: *Cl*imate; *O*rganisms; *R*elief; *P*arent material; *T*ime (81), all of which vary spatially across the landscape and also over time (82) (83) (84). The original parent material of soils is often the underlying bedrock but, due to transportation by natural forces, for many soils the parent rock may be located a significant distance away (17,85). Parent rocks have been broken down through the grinding, compression, expansion and contraction caused by extremes in temperature and the movement of wind, water and ice (20,81). The resulting fragments have since been weathered further through aqueous chemical reactions (86). These processes, in turn, led to an accumulation of minerals above the parent rock, which exhibit enormous diversity according to their location (17) (87) (88) (89) (90). Soil composition develops further chemical complexity through the input of both biogenic inorganic compounds (91) (87), and the organic substances, such as sugars, lipids and biopolymers, generated by plants and animals (92). As plants and microbes grow they alter soil chemistry (20), indeed, microbes are thought to have mediated the processes responsible for the genesis of many minerals (93,94). Organisms affect the characteristics of soil not only through the absorption of nutrients and addition of excreta and decomposition products after death (20), but also by changing its structure, texture and homogeneity through perturbations caused by, for instance, burrowing animals and root growth (4,95).

Figure 2.3 Diagram depicting the process of pedogenesis (96)

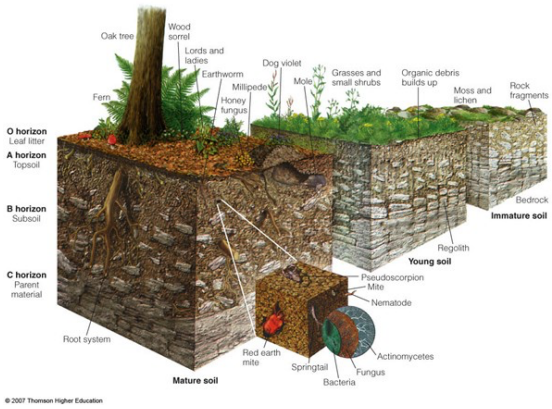


Figure 2.3 The soil forming process

Chemical and mechanical weathering causes the formation of small mineral particulates and the growth, decay and metabolism of living organisms causes a build-up of complex, degraded organic molecules in the upper horizons over geological timescales. Geological forces cause the cycling of minerals into different types of rocks, which exist as facies, which are discrete units of rock with consistent, homogenous characteristics (eg. appearance, mineralogy or fossil content) that are distinct from adjacent units of rock, and the variation chemical composition and depositional environment over different geographical areas forms the basis for the formation of these facies (89) (88).

Geological forces cause the cycling of minerals into different types of rocks, which exist as facies, which are discrete units of rock with comparatively consistent, homogenous characteristics (eg. appearance, mineralogy or fossil content) that are distinct from adjacent units of rock, and the variation chemical composition and depositional environment over different geographical areas forms the basis for the formation of these facies (89) (88). Minerals are generally comprised of compounds containing silicon, oxygen and metals such as aluminium, iron and magnesium, which can be reactive with organic compounds in soil (82) (87) (83) (87). Since mineral surfaces can catalyse some of the reactions affecting organic compounds, one type of mineral may interact with humic materials to produce a different range of organic compounds than the next type of mineral (87) (82) (83) (84), therefore it is possible that the underlying geology of a site may influence the organic chemistry of a site and changes in underlying geology may have an effect on the nature and on the diversity of organic chemicals in soil within sites of forensic interest. Similarly, climactic factors, such as temperature, ambient light, and moisture, as well as biological activity at a specific location vary seasonally throughout the year (87) (84) (83) (82), and since all of these affect the inputs, reaction rate and type of chemical and biochemical reactions occurring in the soil at a

particular location, the variability of the organic profile of soil may be seasonally changeable. As soil horizons develop over time, with more input and cycling of materials, the composition of the soil may develop greater complexity over time, older soils may contain a different range of chemicals than younger soils with the same underlying geology, climate and organisms. All of these factors combined give the potential for the physical, chemical and biological features of earth materials to vary, as distinct, discrete facies, across the surface of the earth, and vertically within soil horizons and geological strata (90) (89) (88).

2.4 Traditional Analytical Techniques for Forensic Geoscience

In light of the variability in types of earth materials, and the important role they play in industries such as mining and engineering, a plethora of analytical techniques have been designed to analyse and compare environmental materials. Since soil is a highly transferable and persistent material that is present almost ubiquitously on land (92) (97) (98), particularly in the secluded areas often encountered in the investigation of serious crimes, it is a potentially a highly informative form of trace forensic evidence. As a result, many of the geological techniques developed for use in academia and industry have been applied to forensic investigations. Despite this, there is a lack of consensus in the published literature concerning how to best classify soils and as a result there are a number of standards, protocols, methods and databases in use (91,17,92).

Of the 4000 recognised types of mineral, most UK soil samples contain six to ten major minerals, in addition to distinctive, rare minerals and, since the major components of most soils are minerals, mineralogical analyses are crucial to the classification of samples (92). Visual examination of particle morphology is also highly informative and some authors maintain that the microscope remains the most powerful tool for forensic geoscience applications (99,7). Many methods of characterisation focus on the bulk physical or chemical properties of soil samples, which are summarised in Table 2.1, however, the use of these bulk techniques can, be problematic when dealing with trace quantities of forensic samples since the sample preparation techniques may be destructive, require homogenisation of the sample, prevent the recovery of the sample after analysis or require large sample amounts.

Table 2.1: Analyses Typically Used in Forensic Geoscience

Analysis Type	Application	Technique	References
Physical Properties	Colour Determination	Standard Colour Charts Reflectance Spectrophotometry	(91,17) (91,100)
	Density	Volumetric Displacement Density Gradient Column	(91,9) (91,9,101)
	Moisture Content	Gravimetric	(17)
	Particle Size Distribution	Sieving XRD Laser Granulometry	(6,102,103) (4,6) (12,104,105)
	Particle Morphology	SEM	(14,15)
	Radio dating		(17)
Chemical Properties	pH		(91,17)
	Conductivity		(91,17)
	Redox Potential		(91,17)
	Ion Chromatography		(91,18,106,107)
	Gas Chromatography		(92) (108) (108) (109) (110) (92) (92) (111) (112)
	High Performance liquid		(18) (30) (31)

	Chromatography		
	Isotope Ratio Mass Spectrometry		(91,113)
	Elemental Analysis	Flame AAS/AES ICP-AAS/AES ICP-MS XRF	(91,114,115) (91,115,116,106,107) (91,115,116,106,107,117) (91,17)
	Mineralogy	Binocular Microscopy Polarised Light Microscopy XRD FTIR SEM	1 1 (17,118) (108,119) (4,91,92,120)
	Isotope Ratio Mass Spectrometry		(91,113)
Biological Properties	Bacterial Profiling Palynology Diatoms		(121) (122) (106) (123) (124) (125)

2.5 Geo-forensic Applications

Due to the variability of soil characteristics over different spatial scales and its transferability, soil has a long history of informing legal investigations (4,12,16,7). In general, geo-forensics can be applied in two distinct scenarios (4,16,18): Firstly, the relatively rare, so-called "seek-and-find" (13) investigations, where assessment of the provenance of a piece of geo-forensic evidence is required; secondly, the more commonly encountered "compare and exclude" (13) problems, where it is necessary to ascertain whether a person or object can be excluded from having made contact with a particular location.

The earliest documented geo-forensic case was a seek-and-find case, which involved a barrel of silver whose contents were found to have been substituted for sand during transportation. Since the route the barrel had travelled was known, the potential location of the substitution could be identified by comparison of the local sand with the questioned sample (4). Similar substitution cases have been reported more recently (17,8). This approach has also been used to refine the search for body deposition sites in no-body murder cases (5,4). In such cases it is necessary to have access to expert knowledge of the geology and land use of the relevant search areas and, even when such resources are available, the investigation may still be unsuccessful (6).

The second type of problem that geo-forensic evidence can be used to help solve, relies on Locard's exchange principle, that every contact leaves a trace (3,126,22,127,128). In these compare-and-exclude cases, relatively small quantities of soil of unknown provenance, which are typically obtained from a suspect or forensically relevant object such as clothing, shoes or a vehicle, are compared with samples known to be associated with a crime (such as the crime scene). During the commission of a crime, contact will be made between perpetrator, crime scene and/or victim and the transfer of material can take place. Thus, trace materials can offer evidence of a contact being made between a suspect and a crime, in an exclusionary comparison and evaluation. This makes soil a particularly valuable form of trace evidence, since its complex composition gives rise to distinctive combinations of features that offer significant variability across geographical locations and regions. This exclusionary approach can be used to corroborate witness testimony (2) or scrutinise an alibi (17,4,129). Furthermore, with the correct interpretation of mixed samples, pre- syn- and post- forensic event activities can be inferred in order to reconstruct crimes (130,106,107,44,131). The geo-forensic community has yet to fully accept and address the implications of mixed source samples in the way evidence is analysed and interpreted, indeed the correct strategy remains the subject of fierce debate (132,130,15,133,116,117).

2.6 Philosophical Considerations

Since locations with poor supervision have been identified as offering the opportunities for criminal activities (134) and soil, in turn, tends to be found in more secluded areas in the UK, it follows that soil is likely to be transferred in a variety of crime scenarios. The clearest example is where a body or exhibit is buried or deposited, indeed many of the case studies presented in the geo-forensic literature involve these types of crime events (4,17,91,8,5,6,135,136,9,11). Given the severity of these crimes, and the consequences of misinterpretation of the evidence, it is essential that reliable methods of comparing "natural" samples of a single, known origin

with “anthropogenic” questioned samples of unknown and possibly mixed provenance are used throughout the sampling, analysis, interpretation and presentation of geo-forensic evidence (137,107,12,13,14,135,15).

The various geological analyses described in the literature are undoubtedly well suited to the analysis of earth materials, and may be used to complement one another to describe such samples in great detail. Unfortunately, the description of a geo-forensic sample, however detailed, cannot always answer all the questions needed to effectively seek-and-find a location or compare-and-exclude two samples as having the same source (12,14,106). Colour and pH, for instance, are aggregate values that are potentially dependent on several components of a sample. Since samples must be homogenised for these procedures, the individual contribution by each component is obscured, and the potential for false positive or false negative interpretation of the data is high in cases involving mixed samples (15,106,20,13,135). A similar difficulty arises when conclusions are based on the relative proportions of different sample components, for instance elemental analysis (15) or particle size analysis (106).

A number of other problematic uses of traditional earth science analytical techniques arise in the published geo-forensic literature, indicating a lack of consideration of the forensic nature of samples. There is no philosophically valid mechanism for obtaining random match probabilities for soil analyses, therefore it is potentially misleading to suggest that the characteristics of a questioned sample “match” those of any other sample (106,14,13,15). Instead, it is preferable to perform additional analyses, of independent properties, to reduce the risk of a coincidental association between samples (137,14). Despite this, mineralogy, colour or elemental analyses, are often presented as corroborative evidence when these qualities are inextricably interdependent (17,138). In addition, many geo-forensic studies have concentrated in the past on comparing samples from sites which may be several miles apart, however, in forensic contexts, the proximity of a crime scene and alibi site may be considerably closer (107). Furthermore, considering that some forensic samples consist of as little as a few grains of material (92), the applicability of some traditional techniques, which require gram (g) quantities of sample, to casework is limited (107,106,15,13,139).

The specific context of a particular crime is of great importance to the correct interpretation of the results of geo-forensic evidence and is therefore essential to the accurate presentation of forensic evidence in court (16). In order for geo-forensic evidence to be of value in court, it must be able to assess the relative likelihood of the evidence given the competing hypotheses presented by the defence and prosecution. These may involve complex explanations for the presence or absence of trace evidence, for instance secondary (or higher level) transfers

coupled with cleaning procedures or decay over time (16,44,131). To ensure correct consideration of the myriad of possible opportunities for the transfer or loss of trace evidence, even in seemingly simple cases, it is necessary to perform experiments which have high degrees of ecological validity. While there have been many studies at the primary level, which build up a body of knowledge regarding initial division and transfer of particulates (140,141,142), further investigations which faithfully reflect the alleged case circumstances are required to ensure the validity of the assumptions made during evidence interpretation (16,131,44).

2.7 Organic Analysis

The organic fraction of soils is mostly biologically derived, but may also contain anthropogenic components, such as agrochemicals and pharmaceuticals (81,31). Organic chemists, like earth scientists, have many tools with which to investigate the composition, structure and reactivity of these substances, some of which are also used in earth science (91,92,143).

Organic chemistry offers a number of advantages in forensic analyses. Due to the sheer number of organic compounds, organic analysis has the potential to be highly discriminatory and is an excellent candidate for using an exclusionary approach in comparing samples. Mixed samples and complex sample matrices, which are typical of forensic cases, are commonly encountered in organic chemistry, therefore there is existing expertise in the separation and purification of mixed samples. Furthermore, since the organic chemistry is not directly linked to the underlying geology of a particular soil, the results of organic analyses can be used to corroborate those of geological analyses (18). In addition, the physical properties of organic compounds are generally different from those of minerals, for instance organic compounds typically have lower melting points and increased solubility. This means that the transfer and persistence of organic compounds may be such that they can be detected in cases where the inorganic fraction of soils cannot (22). Despite this, there have been comparatively few documented attempts to utilise the variability in the organic fraction of soils for forensic purposes, with the notable exception of GC and GC-MS based techniques for the determination of wax markers (108) (109) (110) (92) (92) (111) (112), which has been used successfully in many court cases and is capable of analysing very small traces of soil but is a somewhat complex and time consuming analytical procedure.

2.8 HPLC Analysis

HPLC is an analytical technique specifically designed to separate mixtures of organic compounds on the basis of their polarity, which is a function of their molecular structure (144,145,146). The different types of chemical bonds, and the range and type of the functional groups present in an organic molecule effect the spatial distribution of the electron density within the molecule, and those molecules in which there is a relative concentration of negative and positive electrical charge at different positions in the molecule are referred to as being polar, while molecules in which the negative and positive regions of the molecule are evenly distributed are referred to as being non-polar (147). Due to intramolecular forces such as hydrogen bonding, polar molecules are attracted to other polar substances while non-polar compounds interact most strongly with other non-polar species, non-polar species are hydrophobic and very non-polar substances such as aliphatic hydrocarbons, for instance hexane, are completely immiscible with very polar substances such as water (148).

In HPLC analysis, samples are analysed in the liquid phase, either solid samples dissolved in or liquid samples mixed with a compatible solvent and are typically extracted by shaking or sonicating in a solvent, followed by the removal of any insoluble matter by filtration or centrifugation (144). The liquid sample is then injected onto a column that is densely packed with fine, silica-based particles, known as the stationary phase, and is pumped through the column at high pressure with a solvent, known as the mobile phase (144,146,145). Mobile phases may be pure solvents or mixtures of solvents of different polarity, and the composition of the mobile phase may be altered during the analysis by adding more of a particular solvent or solvents in order to fine tune the polarity or pH of the mobile phase as desired, in addition a variety of stationary phases are available with multiple different types of functional groups bonded to the silica particles, and each of these types of stationary phase has a different polarity. The identity of a molecule is dictated by its structure and composition, which in turn affects its polarity, (147) therefore since each compound interacts slightly differently with both the stationary and mobile phases, each component of the mixture is driven through the HPLC system at a different rate (144,145), causing those compounds of different polarity to elute from the column after a different length of time and different compounds, even stereoisomers, can be separated on the basis of their polarity by HPLC. As the separated compounds elute from the column, they pass through a detector as a series of individual peaks, and the time each constituent spends on the column, known as the retention time, is highly discriminatory (144,146,145), and can be used to identify compounds through the comparison of the retention time obtained with that of a reference standard analysed under the same chromatographic conditions.

When designing an HPLC method for a specific compound, it is necessary to trial different types of stationary phase and mobile phase to tailor the polarity to suit the specific target molecule, in order to ensure that there is sufficient selectivity to the method, that interaction between the stationary phase and the analyte to retain the compound of interest on the column, while also ensuring that the polarity of the mobile phase allows the compound to be eluted from the column at a different time from other related substances known or suspected to be present in the sample (149) (150) (151) (149). It is also necessary to verify the specificity of the method, that any peak detected using a particular combination of mobile phase and stationary phase chemistries relates only to the compound of interest and not to any other co-eluting related substances or unknown substances causing interferences (150) (151) (149). If selectivity and specificity cannot be achieved by altering the mobile phase composition or column chemistry, it is also possible to adjust the flow rate of the mobile phase to alter the speed at which the compounds elute, and therefore calibrate the degree of separation between compounds (149) (150) (151).

There are a number of different types of detector available for use with HPLC, for those which monitor the UV absorbance of the solution eluting from the column at a set wavelength, or a number of different wavelengths simultaneously, or those which record the UV and visible spectra of the solution (149) (150) (151). There are also fluorescence based detectors, and those which monitor the refractive index of the solution, which changes as dissolved compounds elute from the column. In addition, the HPLC may be connected to an electrospray or atmospheric pressure chemical ionisation mass spectrometer in order to record the mass spectrum of the peaks as they exit the HPLC system. In cases where specificity cannot be achieved chromatographically, the choice of detector settings can help to resolve co-eluting peaks of interest since it may be possible to set the detector to record chromatograms at the wavelength or mass-to-charge ratio relating specifically to the compound of interest, and not coeluting compounds (149) (150) (151). Detectors which measure the UV or mass spectra of the peaks also offer a secondary level of identification, as these spectra are complex and highly discriminatory, and the type of detector selected for use during method development is dependent on the properties of the molecule, for instance only certain types of molecule will fluoresce or absorb in the UV-Visible range, while some compounds may be difficult to ionise and therefore unsuitable for MS detection, and it is therefore important to consider the optimum detection method during project planning and method development (144,146,145,149,150,151). Adaptation of the detector settings can also improve sensitivity, which is the ability of the method to detect small amounts of the target analyte, since, for UV detection, the concentration of a substance is related to the absorbance

measured through its extinction coefficient, which varies at different wavelengths, therefore setting the detector to the wavelength at which the target analyte absorbs most strongly allows the analyte to be detected at lower concentrations and for smaller changes in concentration to be detected, thereby improving the precision of the method, however the effect of changing the wavelength must be tested for each compound under each set of chromatographic conditions as any increase in signal strength may counteracted by an increase in background noise, which increases the lower limit of detection (145) (149) (150) (151). Since the detector response is related to the concentration of the sample passing through the detector, it may also be possible to improve sensitivity by increasing the volume of sample injected onto the column, however this may be detrimental to the shape of the peak, which in turn may lead to inaccurate quantification or identification, therefore the optimum injection volume must also be established experimentally during method development for each new compound or new set of HPLC parameters (149) (150) (151). In addition, the linear range of the method varies depending on the analyte and the type of detector used, and it is therefore essential to verify the linearity of the relationship between sample concentration and detector response in order to accurately quantify the compound of interest (152) (151) (150).

Determination of the sensitivity, selectivity, linearity of response and the minimum limits of detection and quantification are essential steps during method development, as these factors are required to validate the ability of the method to reliably identify and quantify the target analyte (61) (58) (60). In addition to developing and validating the HPLC parameters such as flow rate and detector settings, it is also necessary to verify the effect of human and systematic factors related to the preparation of the samples, for instance the relative uncertainty in the weight of sample used, and the number of dilution steps, as these directly effect the accuracy and precision of the technique since systematic errors and uncertainty can accumulate. It is also useful to consider the optimum sample extraction methods in order to ensure that all of the compound of interest can be extracted from typical sample matrices and to determine whether any of the other components typically present in the sample, for instance water, cutting agents or natural compounds such as proteins in biological samples, interfere with the peak of interest in the chromatogram, or inhibit full extraction of the target analyte (60) (152) (61). As discussed in section 2.1.3, it is advisable to incorporate experiments to establish these quality indicators at an early stage in method development, and all of the operational factors relating to sample handling and preparation are also required to validate the ability of newly developed HPLC methods to generate reliable results for that

specific analyte and sample type, in the context of a particular laboratory or organisation, in accordance with ILAC guidelines (61).

HPLC is already widely used in forensic science, particularly for the analysis of drugs of abuse (27) (153) (154) (155) (156) (157), where the identity of the questioned substance, for instance a bag of unidentified tablets, can be ascertained by comparing the retention time of the primary compound in the suspected sample against the retention times obtained for pure, certified reference standards of known drugs of abuse or any control samples of and legitimate compounds offered as a defence proposition, for instance prescription or over the counter drugs. The presence and relative amounts of impurities synthesised during the manufacture, or formed during the degradation of illegal drugs in the HPLC chromatograms of suspected illegal drug samples can be used to exclude different batches of seized narcotics for evidential or intelligence purposes in order to investigate or disrupt organised crime (158) (159). Offence level propositions pertaining to the consumption of illegal drugs, as opposed to simply the possession of illegal drugs, can also be examined by testing samples extracted from bodily fluids such as blood and urine and comparing the chromatograms obtained with those obtained for certified reference standards illegal drugs and their metabolites (160) (28) (161). HPLC can also be used for toxicological purposes (162) (29) (163), for instance to test for poisons in the case of an unexplained death, through comparison of the chromatography obtained from the bodily fluids or tissues of the victim, against reference chromatograms for commonly encountered poisons, or compounds suspected to be relevant to the case, such as prescription medication. Some dyes and inks are organic compounds, or mixtures of different organic compounds and manufacturers use not only different types of compounds, but different formulations to produce the same colour, and therefore the retention times of the constituents of dyes and inks, and the relative size of each peak, may vary between manufacturer, brand or product type and HPLC can be used to discriminate between known and unknown sources for the comparison of questioned documents or coloured fibres (164) (165). In addition, HPLC can be used to identify and quantify explosives and post-blast residues in order to provide intelligence in terrorism cases, for instance by discovering evidence of potential links between terrorist incidents through comparison of the distinctive features of the chromatographic profiles recovered at different scenes (24) (166) (167). The widespread use of HPLC in commercial forensic laboratories, and indeed healthcare, manufacturing and means there is existing capacity and capability to perform HPLC analyses routinely in-house, therefore new HPLC methods can be implemented without any requirement for significant capital expenditure, the procurement of specialist consultant

services, or investment in training or human resources from those forensic science stakeholders for whom best-value is a key priority.

Standard sample preparation techniques for HPLC are analogous to the cleaning stages used in some of the more traditional analyses used in forensic geoscience (8), creating the potential for corroborative, independent analyses to be performed on the same sample, which is ideal in forensic work where sample amounts are often limited. Despite this, the potential of HPLC to analyse geo-forensic samples has only been explored to a limited extent in the published literature. Reuland and Trinler (94) reported the successful discrimination of samples obtained from locations "within a ten mile radius of [Terre Haute] city limits in Vigo County" (Reuland and Trinler, 1981: p204) (31), while similar results were obtained with samples obtained from locations "within the city limits of Alpena, MI, a city of radius 12km" (Siegel and Precord, 1985: p513) (30). partial success was reported in distinguishing rural, urban and semi-rural sites "within a 10-mile radius of Lansing, MI"(Bommarito et. al. 2007: p24) (18) and noted great diversity amongst urban locations.

While these results are promising, the studies suffer from several limitations. Each study compared samples taken from areas of different land use, for instance golf courses, garden lawns and wasteland, which were obtained across large areas (18,31,30). This experimental design does not explore the smaller scale spatial variability of the composition of soils, which limits the potential application of the technique to those cases where sites of interest are situated at locations separated by great distances, which are likely to have differences in the underlying geology and can therefore be discriminated by existing methods, yet does not assist in the many forensic scenarios where it is necessary to discriminate soils within the same site, or a small locality, which are less likely to be able to be discriminated using existing techniques. The previous studies required time consuming and complex sample preparation, which increases the cost of the analysis and introduces potential sources of bias, however as discussed in there is limited demand for costly, labour intensive procedures in the current landscape of commercial forensic service provision in England and Wales due to the prohibitive cost of performing such procedures and since the process of method validation is more challenging, due to the increased difficulty in controlling quality risks makes, for complex procedures. In addition, minimum preparation and analysis times ranged from approximately 150 minutes (min) to overnight (18) (30) (31) (32), and each method detailed seven to ten sample preparation steps, which not only risks delaying the investigation but is also detrimental to the efficiency of the method and productivity of the laboratory and further limits the suitability of applying the technique in profit driven laboratories. Furthermore the minimum sample amount detailed was 1g (31), which may not always be available in forensic

cases, and prevents the use of the technique in cases where only trace amounts of soil are available for comparison.

2.9 Aims of the Current Study

Soil is present at a range of potential crime scenes and the composition of soil is highly variable according to changes in geographical features such as climate, underlying geology, and local plant and animal activity, which allows soil evidence from different locations to be discriminated in order to ascertain provenance through the use of soil databases, or comparative analyses of soils from known and questioned sources. In addition, due to the particulate structure of soil, traces can easily be divided from their source material and transferred to other surfaces, and the measurable properties of soil that facilitate accurate sample comparisons have been demonstrated to be highly persistent in various, forensically relevant scenarios. Soil is therefore valuable for forensic reconstruction, and although there have been a number of studies investigating the potential for using the variability in the inorganic chemical composition of soils at different locations to discriminate samples of forensic interest, there have been relatively few attempts in the published literature to establish independent techniques for forensic soil analysis, for instance through development of methods to analyse the organic chemical composition of the soils. There have been some studies utilising variations in the profiles of wax marker compounds at different locations detected by GC and GC-MS, and untargeted HPLC profiling, however these studies have lacked forensic relevance and ecological validity through the use of unrealistically large sample quantities, impractical or prohibitively expensive analytical procedures, and the spatial resolution afforded by the research design.

In order to develop the previous work and address the gaps in the forensic geoscience literature identified and discussed above, this research aims to:

Develop an HPLC method for the discrimination of close proximity, trace soil and sediment samples for forensic use.

In the interests of achieving this primary aim, a number of objectives will be addressed in order to:

1. Determine whether it is possible to discriminate soil samples obtained from different locations within the same site;

2. Reduce the quantity of soil required to prepare the HPLC sample;

3. Simplify the sample preparation procedure;

4. Reduce the time required to complete each analysis;

5. Apply the technique to samples from locations representing a range of underlying geologies;

6. Monitor temporal variations in soil chromatography;

7. Compare the discriminatory power of the method with existing geoforensic techniques

3 Method Development

3.1 Introduction

Organic molecules can be highly sensitive to certain environmental conditions, and can decompose or degrade when exposed to factors such as moisture, heat, light or other chemicals. As such, the conditions under which soil is stored and prepared has the potential to affect its organic profile. In order for any new analytical method to be considered a useful addition to existing techniques, careful consideration of each step in the sample preparation and analysis process must be assessed to ensure the selection of the optimum methodological parameters that yield the best separation of chromatographic peaks, in the shortest possible time, and that the handling of the sample prior to analysis does not influence the analytical results. In order to verify that the analytical method does not introduce artefacts that could influence the outcome of the comparison, the effects of each step in the laboratory procedure on the profiles obtained was monitored. Furthermore, in order that the method can be satisfactorily commercialised, the cost of consumables, reagents, equipment and analyst time. This chapter documents the preliminary experimental work to develop an improved, less expensive and more user friendly methodology than in previous studies. Many of the studies in this chapter have also been presented by, McCulloch et. al. (33)

3.2 Sampling

Samples were initially collected from Brockwell Park in South London, and an overview of the sample site is shown in figure 3.1. This location was chosen since it represents both potential alibi sites and potential crime scenes. It is well used by the public for a number of different sport and recreational activities, which provides a number of ways in which a person could legitimately come into contact with earth materials, but also contains several areas which could provide opportunities for crimes to be committed, for instance secluded spaces and thoroughfares.

Figure 3.1: Aerial View of Sampling Site used for Method Development Experiments

Figure 3.1 shows the approximate location of the sample site within the UK (in red) and an aerial view of Brockwell Park, and the approximate positions of the individual sampling locations and inter-location distances

Within this site, a purposive sampling approach was followed; four locations were chosen, each with a distinct recreational use or surface vegetation type.

Location 1 was grassland located to the North of the park (Figure 3.2),

Figure 3. 2: Overview of Location 1

Figure 3. 2: Photographs showing (clockwise from top left) an aerial view of location 1, the variation in surface vegetation at location 1, and a close up of the individual sample collection points

Location 2 was wetland located to the West of the park (Figure 3.3).

Figure 3.3: Overview of Location 2

Figure 3. 3: Photographs showing (clockwise from top left) an aerial view of location 2, a close up of the individual sample collection points, and the variation in surface vegetation at location 2

Location 3 was a wooded area located to the South of the park (Figure 3.4)

Figure 3.4: Overview of Location 3

Figure 3.4: Photographs showing (clockwise from top left) an aerial view of location 3, the variation in surface vegetation at location 3, and a close up of the individual sample collection points.

Location 4 was a wild meadow located to the South West of the park (Figure 3.5).

Figure 3.5: Overview of Location 4

Figure 3.5: Photographs showing (clockwise from top left) an aerial view of location 4, the variation in surface vegetation at location 4, and a close up of the individual sample collection points.

Samples were deliberately sought from the organic rich O-Horizon from areas of exposed soil, which were deemed to be more easily transferable and therefore more forensically relevant than deeper horizons or soil covered by dense vegetation. As shown in figures 3.2-3.5, five samples were collected from each location in order to assess intra-location variability, using the grid suggested for sampling footprints and tyre tracks by Pye (91). Approximately 15g was removed from the central point (A) and from points located above, to the right, below, and to the left of the central point (B,C,D and E, respectively). The distance from the centre of point A to the outermost edge of each other point was 50 cm, each sample point was less than 3 cm deep and less than 5 cm in diameter. In accordance with Simmonds (168), samples were gathered using a stainless steel spatula, taking care to remove any turf or gravel, where present.

Samples were stored in individual, uniquely labelled, sealable LDPE bags. In order to prevent cross contamination each sample bag was sealed with adhesive tape and placed in a secondary bag, which was also sealed. Sampling utensils were rinsed with methanol and distilled water then dried with tissue between samples. A representative sample from each location was taken for method development, then each sample was assigned an anonymous identification number by a second analyst in order to minimise cognitive bias of the primary analyst during sample preparation and data analysis, the log cross referencing identification numbers and their sample position was kept confidential throughout sample preparation and initial data analysis. Since organic compounds can be thermally labile and soil temperature is known to influence biological activity (4) (169), samples were stored at -20°C in order to prevent changes to the organic composition of the samples caused by micro-organisms or thermal degradation

3.3 Summary of Method Development Experiments

3.3.1 HPLC Parameters

Each part of the HPLC instrument has the potential to affect the chromatography obtained, therefore careful consideration must be given to the instrument settings used. Column chemistry and mobile phase choice are critical to ensuring that the sample remains on the column long enough to become separated by the differential actions of the mobile and stationary phases, and can be fine-tuned to improve the shape and resolution of peaks. Other factors can also have a significant impact on the observed chromatography, and even changing to another instrument or column of the same make and model can cause unexpected changes to the profiles recorded, therefore it is important to assess the effect of even seemingly

insignificant changes to the analytical method prior to running samples, especially in the absence of reference standards, where like-with-like comparisons are required.

Development of sample preparation and HPLC parameters occurred concurrently therefore different HPLC Parameters were used for some experiments, as described in Table 3.1. All comparisons were like-with-like, however, and made between samples that were analysed with the same HPLC Parameters, during the same run, using the same reagents

Table 3.1 Summary of HPLC Parameters considered during Method Development

HPLC Method Number	Instrument Make/ Model	Column (Manufacturer, Stationary Phase, Length x Internal Diameter)	Mobile Phase Organic Modifier	Gradient	Injection Volume (µl)	Temperature (°C)	
						Sample	Column
A1	Agilent 1200	Zorbax Eclipse, XDB C18, 5µm, 150 x 4.6mm	Acetonitrile	Time (min) / % Organic Modifier: 0.0 / 45; 5.0 / 45; 35.0 / 75; 38.0 / 98; 40.0 / 98; 45.0 / 45; 50.0 / 45	50	5	30
A2	Agilent 1200	Apex II ODS, C18, 5µm, 150 x 4.6mm	Acetonitrile	Time (min) / % Organic Modifier: 0.0 / 45; 5.0 / 45; 35.0 / 75; 38.0 / 98; 40.0 / 98; 45.0 / 45; 50.0 / 45	50	5	30
A3	Agilent 1200	Apex II ODS, C18, 5µm, 150 x 4.6mm	Methanol	Time (min) / % Organic Modifier: 0.0 / 45; 5.0 / 45; 35.0 / 75; 38.0 / 98; 40.0 / 98; 45.0 / 45; 50.0 / 45	50	5	30
A4	Agilent 1200	Apex II ODS, C18, 5µm, 150 x 4.6mm	Methanol	Time (min) / % Organic Modifier: 0.0 / 5; 95.0 / 95	50	5	30

				97.0	5			
				105.0	5			
A5	Agilent 1200	Waters Symmetry C18, 3.5 µm, 150 x 4.6mm	Acetonitrile	Time (min)	% Organic Modifier	50	5	30
				0.0	45			
				5.0	45			
				35.0	75			
				38.0	98			
				40.0	98			
				45.0	45			
				50.0	45			
A6	Agilent 1200	Waters, X-Bridge Phenyl, 3.5 µm, 100 x 4.6mm	Acetonitrile	Time (min)	% Organic Modifier	50	5	30
				0.0	45			
				5.0	45			
				35.0	75			
				38.0	98			
				40.0	98			
				45.0	45			
				50.0	45			
A7	Agilent 1200	Waters, X-Bridge C18, 3.5 µm, 100 x 4.6mm	Acetonitrile	Time (min)	% Organic Modifier	50	5	30
				0.0	45			
				5.0	45			
				35.0	75			
				38.0	98			
				40.0	98			
				45.0	45			
				50.0	45			
A8	Agilent 1200	Waters, X-Bridge C18, 3.5 µm, 100 x 4.6mm	Methanol	Time (min)	% Organic Modifier	50	5	30
				0.0	5			
				95.0	95			
				97.0	5			
				105.0	5			
A9	Agilent 1200	Waters, X-Bridge C18, 3.5 µm, 100 x 4.6mm	Methanol	Time (min)	% Organic Modifier	50	5	30
				0.0	45			
				5.0	45			
				35.0	75			

				38.0	98			
				40.0	98			
				45.0	45			
				50.0	45			
A10	Agilent 1200	Waters, X-Bridge C18, 3.5 µm, 100 x 4.6mm	Ethanol	Time (min)	% Organic Modifier	50	5	30
				0.0	5			
				95.0	95			
				97.0	5			
				105.0	5			
A11	Agilent 1200	Waters, X-Bridge C18, 3.5 µm, 100 x 4.6mm	Ethanol	Time (min)	% Organic Modifier	50	5	30
				0.0	45			
				5.0	45			
				35.0	75			
				38.0	98			
				40.0	98			
				45.0	45			
				50.0	45			
A12	Agilent 1200	Waters, X-Bridge C18, 3.5 µm, 100 x 4.6mm	Acetonitrile	Time (min)	% Organic Modifier	50	5	30
				0.0	40			
				5.0	50			
				30.0	75			
				31.5	98			
				33.5	98			
				35.0	40			
				37.0	40			
A13	Agilent 1200	Waters, X-Bridge C18, 3.5 µm, 100 x 4.6mm	Acetonitrile	Time (min)	% Organic Modifier	50	5	30
				0.0	47			
				28.0	75			
				29.0	98			
				31.0	98			
				31.1	47			
				35.0	47			
A14	Agilent 1200	Waters, X-Bridge C18, 3.5 µm,	Acetonitrile	Time (min)	% Organic Modifier	50	5	30
				0.0	47			
				28.0	75			

		100 x 4.6mm		29.0	98			
				31.0	98			
				32.0	47			
				35.0	47			
A15	Agilent 1200	Waters, Symmetry C18, 3.5 µm, 150 X 4.6mm	Acetonitrile	Time (min)	% Organic Modifier	50	5	30
				0.0	47			
				28.0	75			
				29.0	98			
				31.0	98			
				32.0	47			
				35.0	47			
A16	Agilent 1200	Waters, X-Bridge C8, 3.5 µm, 100 x 4.6mm	Acetonitrile	Time (min)	% Organic Modifier	50	5	30
				0.0	47			
				28.0	75			
				29.0	98			
				31.0	98			
				32.0	47			
				35.0	47			
A17	Agilent 1200	Fortis, H2O, 5 µm, 150 X 4.6mm	Acetonitrile	Time (min)	% Organic Modifier	50	5	30
				0.0	47			
				28.0	75			
				29.0	98			
				31.0	98			
				32.0	47			
				35.0	47			
A18	Agilent 1200	Fortis, H2O, 5 µm, 150 X 4.6mm	Acetonitrile	Time (min)	% Organic Modifier	50	5	30
				0.0	5			
				95.0	95			
				97.0	5			
				105.0	5			
A19	Agilent 1200	Waters, X-Bridge C8, 3.5 µm, 100 x 4.6mm	Acetonitrile	Time (min)	% Organic Modifier	50	5	30
				0.0	40			
				5.0	50			
				30.0	75			
				31.5	98			

				33.5	98			
				35.0	40			
A20	Agilent 1200	Waters, X-Bridge C8, 3.5 µm, 100 x 4.6mm	Acetonitrile	Time (min)	% Organic Modifier	50	5	30
				0.0	42			
				3.0	50			
				25.0	65			
				27.3	98			
				29.5	98			
				30.5	42			
				35.0	42			
A21	Agilent 1200	Waters, X-Bridge C8, 3.5 µm, 100 x 4.6mm	Acetonitrile	Time (min)	% Organic Modifier	50	5	30
				0.0	47			
				25.0	60			
				27.3	98			
				29.5	98			
				30.5	42			
				35.0	42			
A22	Agilent 1200	Waters, X-Bridge C8, 3.5 µm, 100 x 4.6mm	Acetonitrile	Time (min)	% Organic Modifier	50	5	30
				0.0	47			
				5.0	55			
				26.0	74			
				31.0	98			
				32.0	47			
				35.0	47			
A23	Agilent 1200	Waters, X-Bridge C18, 3.5 µm, 100 x 4.6mm	Acetonitrile +0.1% Formic Acid	Time (min)	% Organic Modifier	50	5	30
				0.0	47			
				5.0	55			
				26.0	74			
				31.0	98			
				32.0	47			
				35.0	47			
A24	Agilent 1200	Waters, X-Bridge C18, 3.5 µm, 100 x 4.6mm	Acetonitrile	Time (min)	% Organic Modifier	50	Ambient	30
				0.0	47			
				3.0	55			
				24.0	74			

				29.0	98			
				31.0	98			
				32.0	47			
				35.0	47			
A25	Agilent 1200	Waters, X-Bridge C18, 3.5 µm, 100 x 4.6mm	Acetonitrile	Time (min)	% Organic Modifier	50	5	50
				0.0	47			
				3.0	55			
				24.0	74			
				29.0	98			
				31.0	98			
				32.0	47			
				35.0	47			
A26	Agilent 1200	Waters, X-Bridge C18, 3.5 µm, 100 x 4.6mm	Acetonitrile	Time (min)	% Organic Modifier	50	5	30
				0.0	47			
				3.0	55			
				24.0	74			
				29.0	98			
				31.0	98			
				32.0	47			
				35.0	47			
A27	Agilent 1200	Waters, X-Bridge C18, 3.5 µm, 100 x 4.6mm	Acetonitrile +0.1% Formic Acid	Time (min)	% Organic Modifier	50	5	30
				0.0	47			
				3.0	55			
				24.0	74			
				29.0	98			
				31.0	98			
				32.0	47			
				35.0	47			
S1	Shimadzu VP	Whatman Partasil C18, 10µm, 250 x 4.6mm	Acetonitrile	Time (min)	% Organic Modifier	10	10	25
				0.0	65			
				100	65			
S2	Shimadzu VP	Whatman Partasil C18, 10µm,	Acetonitrile	Time (min)	% Organic Modifier	10	10	25
				0.0	5			
				90	95			

		250 x 4.6mm		95	95			
S3	Shimadzu VP	Whatman Partasil C18, 10µm, 250 x 4.6mm	Acetonitrile	Time (min)	% Organic Modifier	20	10	25
				0.0	30			
				10.0	30			
				60.0	80			
				65.0	80			
				65.1	98			
				70.0	98			
S4	Shimadzu VP	Whatman Partasil C18, 10µm, 250 x 4.6mm	Acetonitrile	Time (min)	% Organic Modifier	20	10	25
				0.0	40			
				10.0	40			
				15.0	50			
				31.0	50			
				50.0	80			
				55.0	80			
				55.1	98			
				60	98			
S5	Shimadzu VP	Whatman Partasil C18, 10µm, 250 x 4.6mm	Acetonitrile	Time (min)	% Organic Modifier	20	10	25
				0.0	45			
				5.0	45			
				35.0	75			
				38.0	98			
				40.0	98			
S6	Shimadzu VP	Whatman Partasil C18, 10µm, 250 x 4.6mm	Acetonitrile	Time (min)	% Organic Modifier	50	10	25
				0.0	45			
				5.0	45			
				35.0	75			
				38.0	98			
				40.0	98			
S7	Shimadzu VP	Whatman Partasil C18, 10µm, 250 x 4.6mm	Acetonitrile	Time (min)	% Organic Modifier	50	10	25
				0.0	45			
				5.0	45			
				35.0	75			

<table>
<tr><td></td><td></td><td></td><td></td><td><table><tr><td>38.0</td><td>98</td></tr><tr><td>45.0</td><td>98</td></tr></table></td><td></td><td></td><td></td></tr>
<tr><td>S8</td><td>Shimadzu VP</td><td>Whatman Partasil C18, 10µm, 250 x 4.6mm</td><td>Acetonitrile</td><td><table><tr><td>Time (min)</td><td>% Organic Modifier</td></tr><tr><td>0.0</td><td>45</td></tr><tr><td>5.0</td><td>45</td></tr><tr><td>35.0</td><td>75</td></tr><tr><td>38.0</td><td>98</td></tr><tr><td>40.0</td><td>98</td></tr><tr><td>45.0</td><td>47</td></tr><tr><td>50.0</td><td>47</td></tr></table></td><td>50</td><td>10</td><td>25</td></tr>
</table>

3.3.1.1 Replication of Bommarito et. al. (17)

In a chromatographic separation, it is desirable to achieve the best resolution of peaks in the shortest time possible. Of the previous studies reported in the literature, Bommarito et. al. (18) presented the best chromatography, with better resolution of peaks and higher signal to noise ratios than the previous studies, enabling unambiguous assignment of distinctive peaks, however the instrument parameters necessitated an unusually long run time, which is highly detrimental to the practicability of the method. Initial chromatographic conditions were therefore based on this paper with the aim of optimising the conditions to reduce the analysis time. This work is also discussed in McCulloch et. al. (33)

Method

In order to verify the results of Bommarito et. al. (18), the original instrument parameters were adhered to as faithfully as was practicable, however it was necessary to use an alternative type of HPLC and column manufacturer, as the original column was an older model that was not available for general purpose in the UK, and therefore was not considered a practical option for routine analysis in the UK. The system was a Shimadzu VP series, which was comprised of an SIL-10AD autosampler with cooling tray, a FCV-10AL solvent mixing system, a DGU-14A vacuum degasser, a SCL-10A control unit, a SPD-M10A diode array detector (DAD) and a CTO-10AS column oven. The column used was a Whatman Partasil 10 ODS analytical column (10um particle size, C18 packing material, column dimensions 250 x 4.6mm). In addition, since they are not detailed in the original paper, standard values were used for the DAD and column temperature settings.

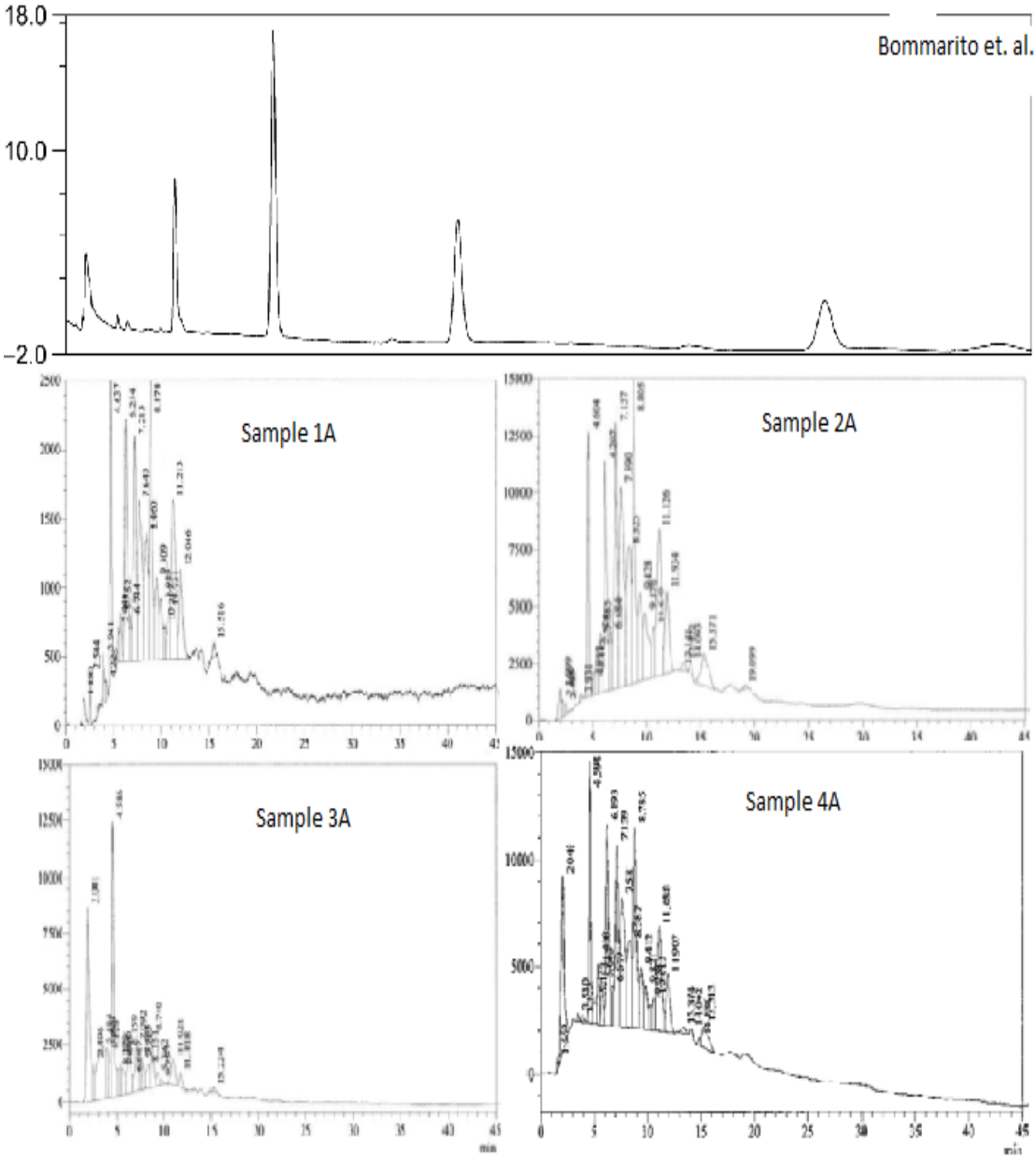
HPLC samples were prepared from points 1A, 2A, 3A, and 4A, approximately 1g was weighed into a centrifuge tube and 1ml acetonitrile was added by pipette. The samples were then mixed and sonicated for 20 min at ambient temperature then centrifuged at 10,000 rpm for 10 min. The supernatant sample solution was transferred to an HPLC vial for analysis and analysed HPLC Method S1 (see Table 3.1).

Results and Discussion

The chromatography obtained by Bommarito et. al. (18) could not be replicated for this study (figure 3.6), (see also McCulloch et. al. (33)). The results obtained in this study exhibited many more peaks, each of which exhibited poor retention and resolution with respect to Bommarito et.al. (18) and all peaks eluted much earlier in the chromatograms than for the Bommarito et. al. (17) study.

As presented by McCulloch et. al. (33) it was not possible to replicate the chromatography obtained in the Bommarito et. al. (18) study, which presented chromatograms containing as few as six well resolved peaks eluting over the entire 100min run and which, in turn, allowed easy visual comparison of results. In contrast, all of the samples taken from Brockwell Park exhibited poor chromatography with between 20 and 25 peaks. All peaks eluted within 20mins, and the vast majority of peaks in this study were not baseline resolved (Figure 3.6).

Figure 3.6: Chromatography generated by replication of Bommarito et. al. (18) instrument parameters for each sample location



Poor resolution is problematic in two ways; not only is there potential for the presence of multiple components with similar retention times to be masked, but it is also difficult to accurately identify peaks with similar retention times, due to the precision limits of the

instrument. It was, therefore, necessary to redevelop the method with the additional aim of improving peak resolution. The poor resolution obtained in this study is likely to be the result of the use of a slightly different column from the original study, the columns were of the same dimensions and contained C18 irregular shaped silica stationary phase in both experiments, however Bommarito et. al. (18) used a 5um particle size, which typically results in improved peak resolution, and a different column manufacturer, which can also affect performance. Different types of columns were therefore tested in future experiments (section 3.3.1.2)

3.3.1.2 Column

The choice of column is one of the most important steps in HPLC method development, since sample components must have some affinity to the stationary phase in order to be retained and separated from one another. The ligands attached to the silica particles are the most important factor dictating sample retention, with longer aliphatic chains having a higher affinity for, and therefore stronger retention of less polar sample components. The quality and size of silica used in column manufacturing, and the chemistry used to bind the ligands to the silica also has a significant effect on the peak shape obtained, and therefore the efficiency of the separation. Smaller particle sizes maximise the stationary phase surface area, allowing enhanced interaction between the sample and stationary phase, thereby improving the separation possible in comparison with columns of the same length. Smaller particles therefore allow shorter columns to be used, which means peaks elute more quickly, reducing the run time, and since peak broadening due to partitioning between the mobile and stationary phases is directly proportional to the time spent on the column, peak shape and resolution is also improved by reducing particle size.

Method

There are a great many different HPLC columns available commercially, however, it is not practical in any method development project to trial even a fraction of the columns on offer, due to their cost. Initial analyses were performed using a Whatman Partasil C18 column, with 10μm particle size, which was 250mm long and had an internal diameter of 4.6mm, on HPLC samples prepared from well mixed soil, prepared as per section 3.3.1.1 using method S8. Upon transferring the project to a new laboratory, this chromatography was compared with that obtained for a Zorbax Eclipse XDB C18 column and an Apex II ODS C18 column, which were both 150mm long with a 4.6mm internal diameter and 5μm particle size for samples, using methods A1 and A2, respectively (table 3.1).

More modern columns were subsequently tested, with 3. 5μm particle size and a range of stationary phases and the chromatography achieved was compared with previous results. The

columns used were Waters Symmetry C18 (3. 5µm, 150 x 4.6mm), Waters X-Bridge Phenyl (3. 5µm, 100 x 4.6mm), Waters X-Bridge C18 (100 x 4.6mm), Waters X-Bridge C8(100 x 4.6mm), Fortis H2O (5 µm, 150 X 4.6mm) using methods A5-7, A16 and A17, respectively (table 3.1).

Results and Discussion

Figure 3.7 shows that most of the peaks elute towards the end of the run using these longer C18 based columns and very few peaks elute in the first 10 minutes, suggesting that use of a shorter column would be advisable to reduce run times and improve productivity, however the best separation was obtained for the Apex II ODS column.

Figure 3.7: Chromatograms obtained for Whatman Partasil (top), Apex II ODS (middle), and Eclipse XDB C18 (bottom) columns

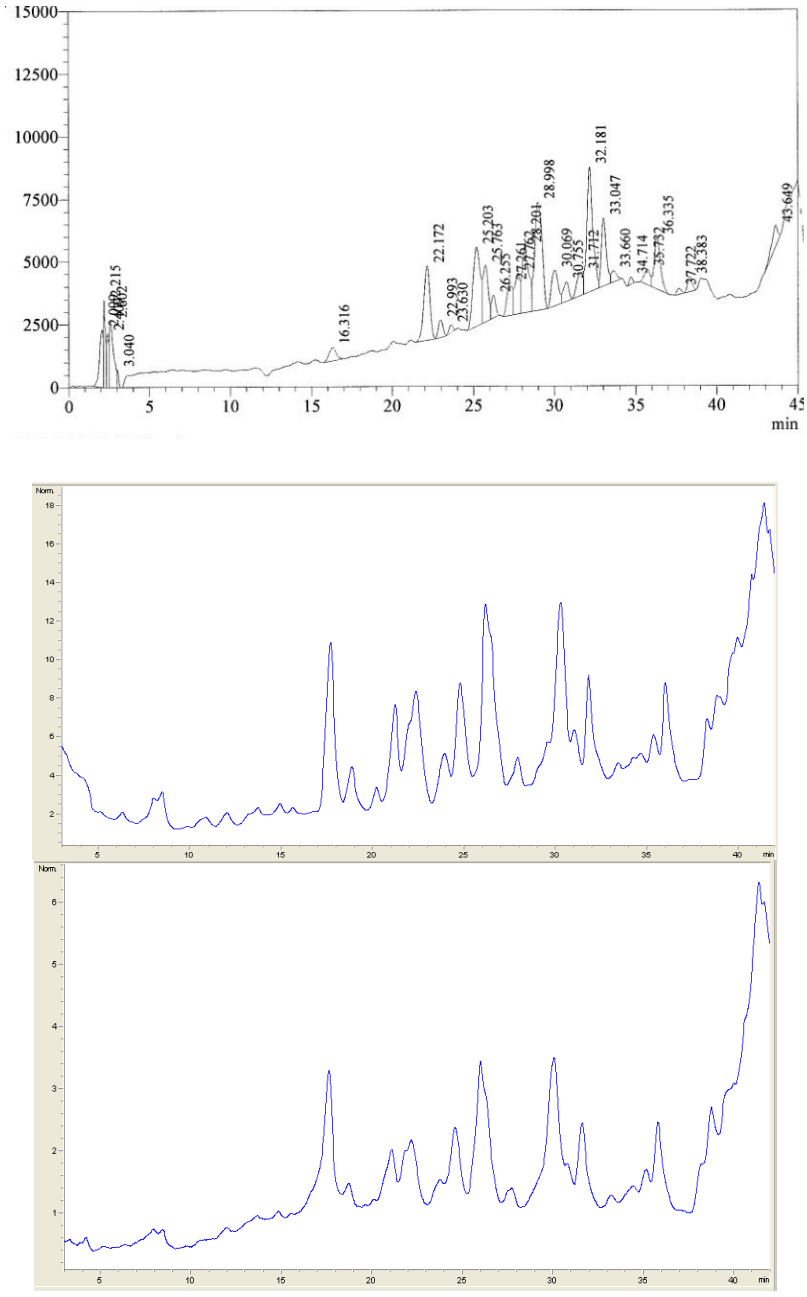


Figure 3.8 shows that the retention times were shortest for the Xbridge C8 column (shown in green), however the peak shape was unsatisfactory and resolution of peaks was poor. Both of the columns with C18 Column chemistry, the Fortis H_2O (shown in blue) and the Waters Symmetry C18 (shown in red) retained the sample components on the column effectively and produced chromatograms with good peak shape, however the retention times were still longer than required, due to the use of the 150mm column.

*Figure 3.8: Chromatograms obtained using Fortis H2O 150*4.6, 5µm (Blue), Symmetry C18 150*4.6 (Red), 5µm, and Xbridge C8 100*4.6, 3.5µm (Green) columns*

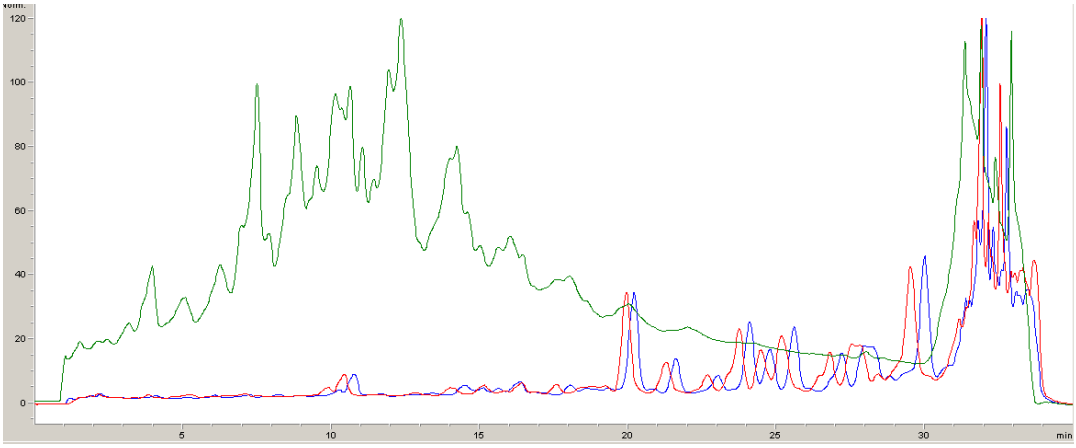
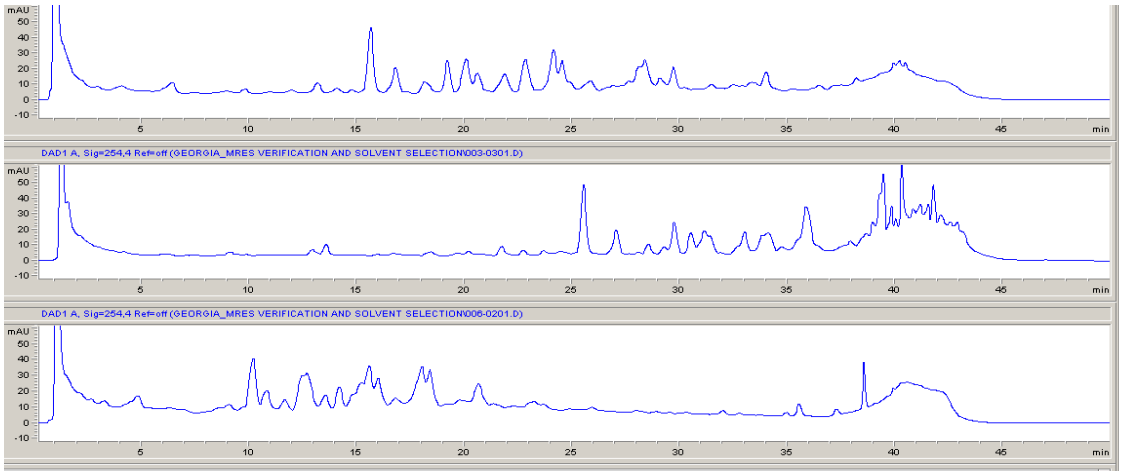


Figure 3.9 shows that both the C18 (top and bottom chromatograms) and the Phenyl (middle chromatogram) retained the compounds of interest, and gave reasonable peak shape and resolution, however the Phenyl column retained the peaks for longer than required, which reduces the efficiency of the method as fewer samples can be analysed in any given period of time. The length of time a compound stays on the column increases the variability in retention time for each compound, widening the retention time window and causing the peaks to broaden. Since the aim of any chromatographic technique is to separate compounds, broad peaks are problematic as this can cause coelution of compounds, leading to inaccruate identification and quantification. C18 was selected as the most appropriate column chemistry since the peaks were of a good shape and well separated throughout the duration of the run.

*Figure 3.9: Symmetry C18 150*4.6, 5µm (Top); Xbridge Phenyl 100*4.6, 3.5µm (Middle); Xbridge C18 100*4.6, 3.5µm (Bottom)*



Use of a 100mm column (bottom chromatogram) shortened the retention times of the compounds of interest compared to the 150mm column (top chromatogram) and was therefore selected for use, however it was recognised that it would be necessary to adjust the mobile phase gradient to optimise the resolution of the peaks of interest.

3.3.1.3 Mobile Phase

The relative polarity of the mobile and stationary phase environment dictates the resolution and peak shape of sample components. Slight differences in the organic content, and the polarity of that solvent can have a pronounced effect on the chromatography recorded and the particular separation required for a specific task, for instance isolation of a particular isomer or pair of related substances, can be fine-tuned through balancing the ability of the column to retain, with the ability of the mobile phase to elute compounds of interest. A range of solvents suitable for reverse phase HPLC, with varying levels of polarity were used as sample diluent and mobile phase organic modifier.

Method

An HPLC sample was first prepared as per section 3.3.1 and analysed using HPLC method A7 (table 3.1). Another HPLC sample was prepared using methanol in place of acetonitrile, and analysed using HPLC methods A3, A4, A8 and A9 (table 3.1). A third HPLC sample was prepared using ethanol in place of acetonitrile, and analysed using HPLC methods A10 and A11 (table 3.1).

Peak shape is sensitive to changes in pH, since molecular and ionised forms of the same compound have slightly different distributions of electron density, and therefore have slightly different polarities and retention times. In general, it is recommended to adjust mobile phase

pH to beyond 2pH units from a sample's pKa (the pH at which the compound is equally distributed between ionised and molecular form). In order to assess the effect of mobile phase acidification on peak shape, samples previously analysed with acetonitrile and water, were re-analysed with the solvents containing 0.1% Formic acid, see HPLC methods A23 and A27 (table 3.1)

Results and Discussion

Figure 3.10 shows that both methanol (shown in green) and acetonitrile (shown in blue) allowed a large number of peaks to be detected with strong signal to noise ratios, whereas there were fewer peaks detected using 50:50 acetonitrile water (red) and ethanol (pink) and those that were detected had relatively poor signal to noise ratios. Since the methanol gave rise to longer retention times and the resolution and peak shape were preferable using acetonitrile, acetonitrile was chosen as the best organic modifier.

Figure 3.10: Chromatograms obtained using Methanol (Green) Ethanol (Pink) Acetonitrile (Blue) and 50:50 Water: Acetonitrile (Red) as the organic modifier

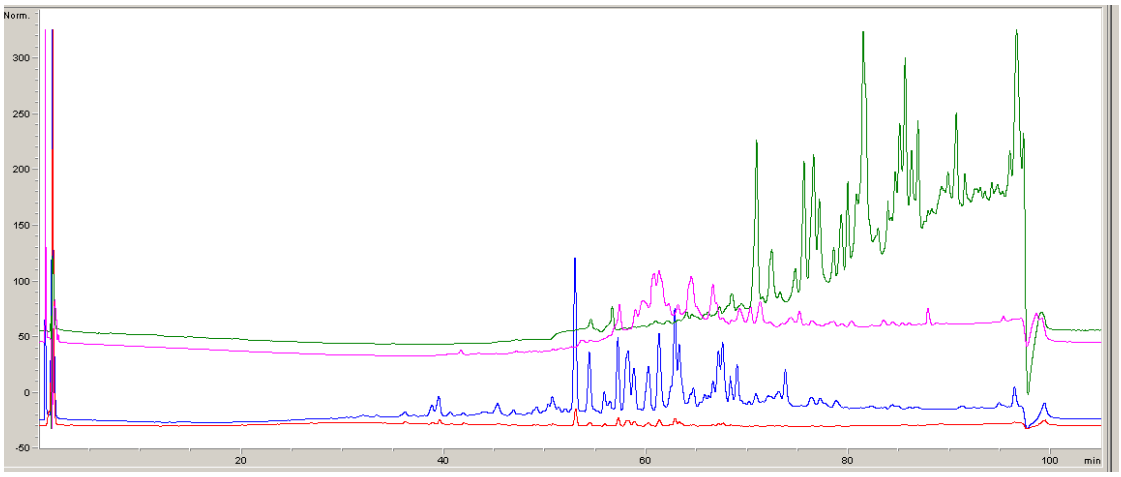
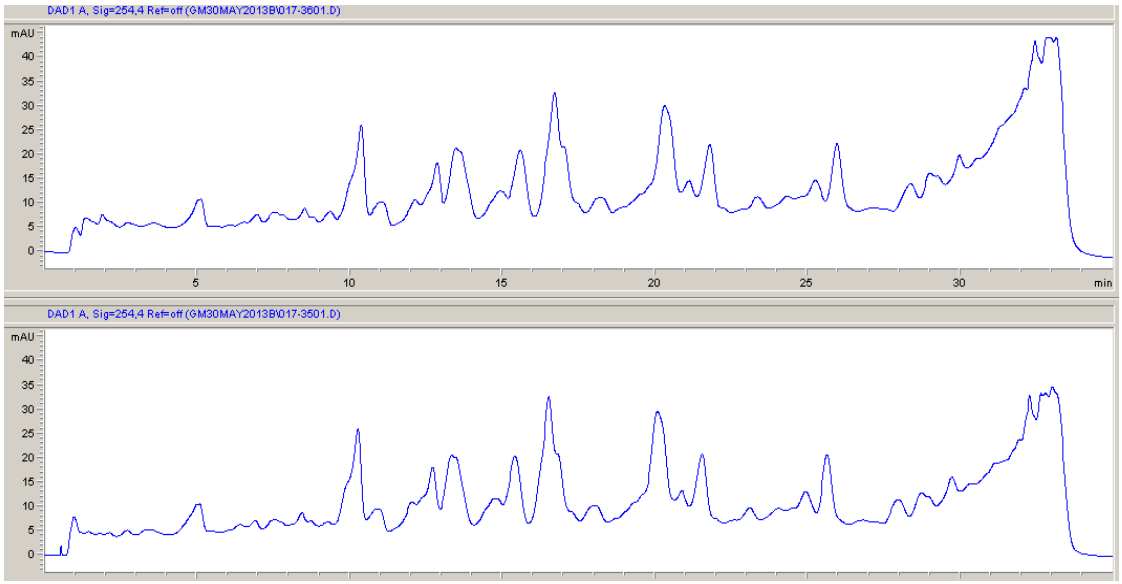


Figure 3.11 shows that there was no improvement to the peak shape or resolution upon the addition of 0.1% formic acid to the mobile phase and diluent, therefore, since the column stationary phase is susceptible to degradation under acidic conditions, no acid was used in the remaining experiments.

Figure 3.11 Chromatography obtained using solvents containing 0.1% Formic Acid (top) and solvents without formic acid (bottom)



3.3.1.4 Gradient

The retention time of late eluting compounds, which have higher affinity for the organic ligands attached to the stationary phase, can be reduced using gradient elution, where the organic content of the mobile phase is gradually increased throughout the run, rather than isocratic elution, where the mobile phase composition remains constant. In order to determine the optimal conditions for separation of the components present in the sample, a series of mobile phase gradients were used, and are detailed in Table 3.1. The aim of each successive change to the gradient was to separate the mixture into the greatest number of components, with the maximum possible separation between each peak, in a reasonable, practical run time.

Method

A sample was prepared as per section 3.3.1.1 and injected using HPLC Methods S2-S8 (table 3.1) and these experiments are also presented by McCulloch et. al (33). Upon transferring the study to another laboratory, with different columns and a different make of HPLC system, additional gradient development was performed. New samples were prepared as per 3.3.1.1, then analysed with gradients A7, A12 and A24 (table 3.1).

Results and Discussion

The successive changes to the mobile phase composition presented by McCulloch et. al (33) and detailed in Figure 3.12 (Table 3.1, gradients S2-S5) reduced run time, improved the shape of the baseline and separation of peaks, revealing additional sample components. Peaks were carried over from one sample to the next, therefore a wash step was added, however this

caused interferences in the subsequent chromatogram (Table 3.1, gradient S7) which were removed by the addition of a column conditioning step (Table 3.1, gradient S8)

Figure 3.12 Chromatograms obtained following successive changes to the gradient using methods S2-S8 (top to bottom)

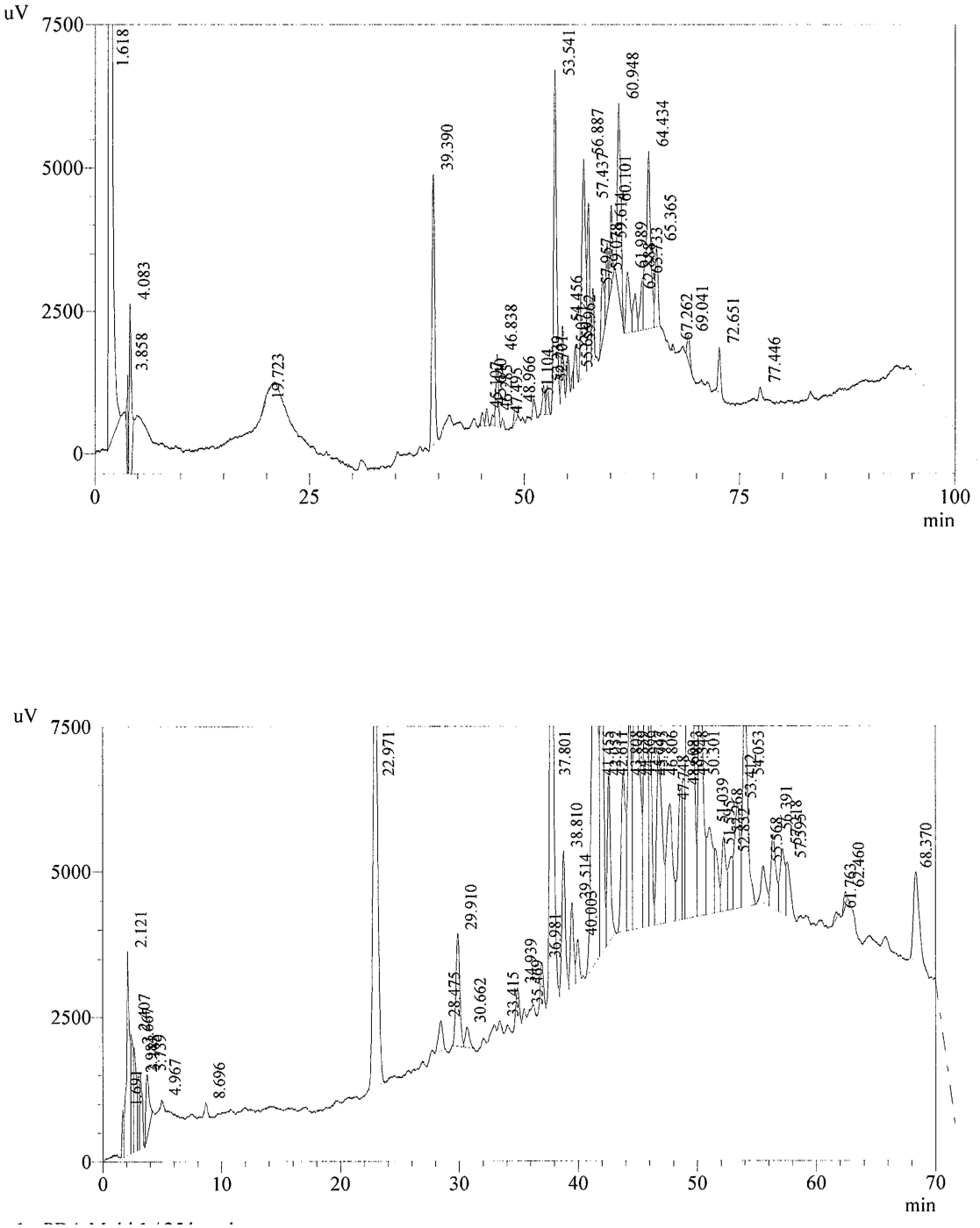


Figure 3.12 Continued

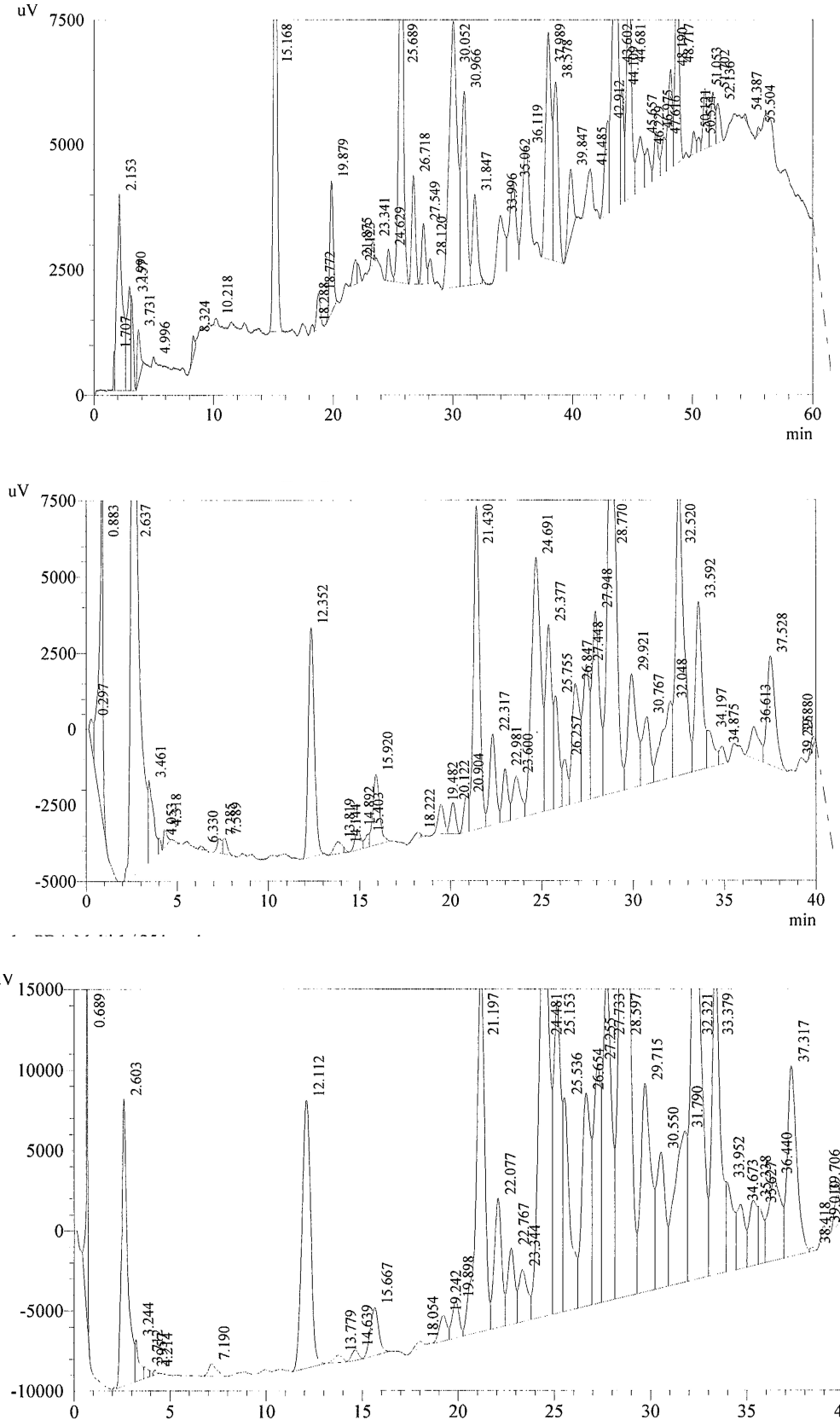
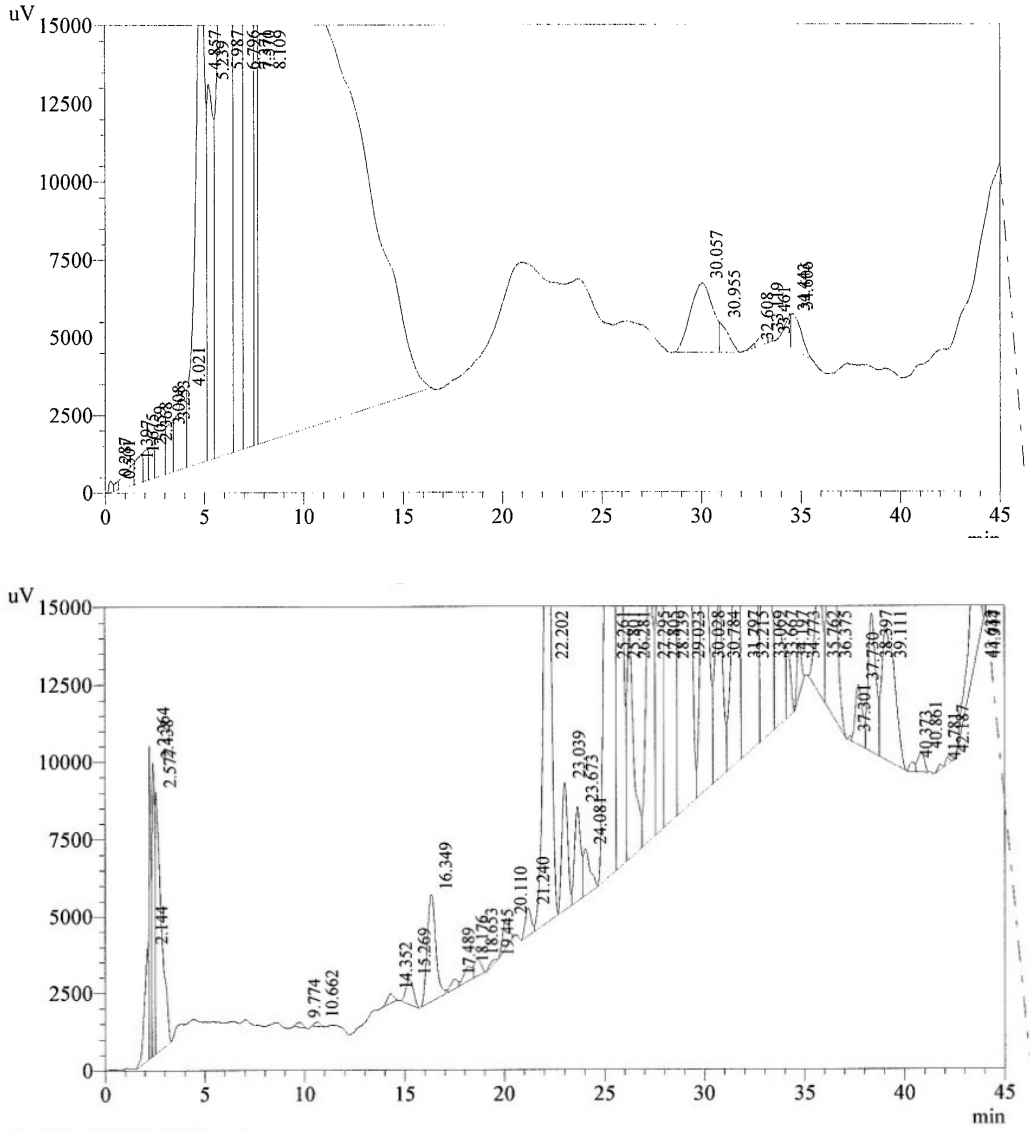


Figure 3.12 Continued



The use of gradient elution achieved better separation of the components of the samples, with respect to that obtained during the replication of the method conditions of Bommarito et.al. (18) (see section 3.3.1.1, figure 3.6) This allowed more peaks to be detected, thereby offering more points for comparison between samples and aiding the assignment of peak identities.

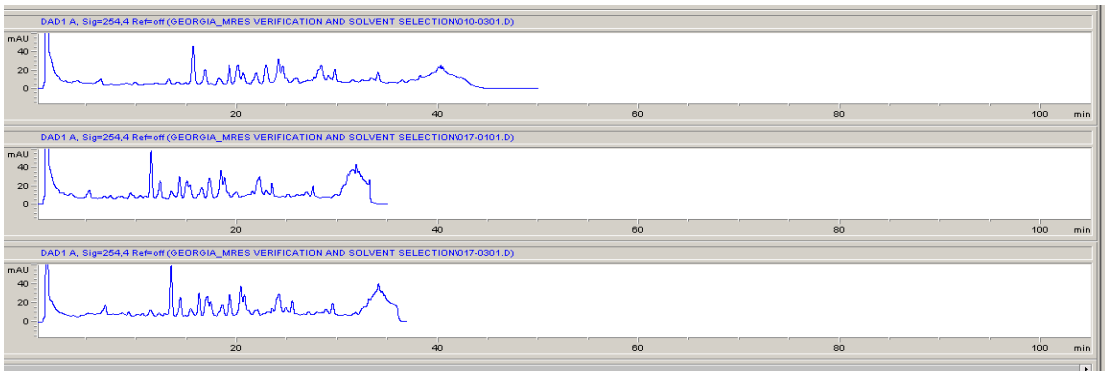
As can be seen in Figure 3.12, in Gradient S2 most peaks elute after 40min, when the gradient reaches around 45% MeCN, and before 75 min, where the mobile phase is approximately 80% MeCN. As such, Gradient S3 was held for 10 min at 30% MeCN, to maintain retention of early eluting polar compounds, and then increased over the optimum range identified in the previous run to 80% MeCN. In order to ensure highly non-polar compounds were completely removed from the column, the mobile phase was raised to 98%. Gradient S3 generated much better separation since the peaks eluted over a longer section of the run, which allowed an additional 11 peaks to be detected. Since the majority of peaks eluted between 50-80%

MeCN, the starting composition of Gradient S4 was raised to 40% MeCN to retain early eluting compounds, then quickly raised to 50% MeCN to reduce the run time, then slowly increased to 80% to ensure suitable resolution of peaks, and finally washed with 98% MeCN. Gradient S4 yielded better separation than Gradient S3 (Figure 3.12), however the two stage gradient created an unstable baseline which made peak identification more difficult. It is also more difficult to achieve good repeatability with complex gradients due to slight differences in pumping efficiency. Gradient S5 was simplified to a single gradient over 30min, raising the starting composition to 45% MeCN and reducing the hold time in order to further reduce the run time.

An unusual baseline and large peaks eluting at the start of the run were observed, which indicated that the wash step was insufficient, therefore an additional 5 minutes were added in Gradient S7 to ensure no peaks were carried over from one injection to the next. This caused a large interference peak to elute over the first 15min of the subsequent chromatogram, which was thought to be the result of the large, sudden change in mobile phase composition (Figure 3.12). An additional 5 min was, therefore, added to allow the column to equilibrate, giving a total final run time of 50min for gradient S8.

Shortening the chromatographic run time is non-trivial when one considers that MeCN has been in short supply in recent years, causing project delays and price increases. The reduction in run time achieved in this project not only doubled potential productivity, but also reduced the MeCN cost by £1.57 per run. The 100min run time detailed by Bommarito et. al (18) is severely detrimental to the efficiency of the analysis and, likewise, to the productivity of any laboratory performing the analysis. In order to ensure correct functioning of the instrument and accurate sample preparation it is necessary to run a number of system suitability injections, furthermore a series of standards must be injected if quantitative analyses are to be performed (15) (116). Such tests are mandatory in a regulated environment like a forensic laboratory. A 100min run time would permit only 14 injections to be performed per day, which means an entire day could be required for analysing standards and system suitability solutions prior to sample analysis. This is clearly detrimental to the practicability of the method. Gradient S8 was therefore selected for use in the preliminary feasibility study (Chapter 4) to determine whether the variability of the HPLC profiles of soil was sufficient to allow for the discrimination of close proximity samples.

Figure 3.13 Chromatograms obtained following successive changes to the gradient using methods A7 (top) A24 (middle) and A12 (bottom)



Following the feasibility study, the development of the mobile phase gradient continued on the Agilent instrument, using new columns, to further improve the sample preparation and analysis method, and the results in Figure 3.13 show successive changes in chromatography for the Xbridge C18 100*4.6, 3.5µm column, using the gradients shown in Table 3.1 for A7 (top) A24 (middle) and A12 (bottom). These chromatograms show that it is possible to further reduce the run time to 35min without significantly compromising the peak shape, resolution, or the number of peaks observed. Gradient A24 was therefore selected for use in the subsequent studies into the development of data analysis strategies (Chapter 5), the temporal and geographic variability of the profiles (Chapter 6), and the comparison of HPLC with more established analytical techniques for forensic geoscience (Chapter 7).

3.3.1.5 Injection Volume

Peak size is proportional to the amount of sample injected onto the column, therefore sample concentration can often be reduced without a loss of signal by increasing the injection volume, however these effects may be non-linear and there can be a loss of resolution due to peak broadening resulting from larger injection volumes. Since sample quantity is a key limiting factor in forensic investigations, the injection volume was successively increased to allow the amount of sample required to be reduced without affecting the sensitivity of the method.

Method

A sample of well-mixed soil from point 4A was prepared according to the method described in section 3.3.1.1 and injected at 10µl and 20µl using HPLC Methods S2 and S3, then the same soil was re-prepared according to the method described in 3.1.1.3 and injected at 50µl using HPLC Method S6.

Results and Discussion

Increasing the volume of sample injected onto the system from 10µl to 20µl increased the height of the peaks without significantly affecting peak resolution (Figure 3.12, S2 and S3 respectively). For instance, in Gradient S2 the main peak at 39.390min increases in height from 4732µV to 8048µV in Gradient S3 (equivalent peak at 22.971min). This allowed the sample concentration, and consequently the amount of sample required for analysis, to be reduced without losing sensitivity. It should be noted, however, that the peak responses were generally very weak, a typical target response in HPLC is between 0.5V and 1V. Despite this, the limit of detection, defined as the peak response with a signal to noise ratio of at least 3:1 (152), was calculated to be as low as 174µV. To compensate for very low peak responses, the injection volume was further increased from 20µl to 50µl in Gradient S6, which increased the response of the characteristic peak eluting at 12min from 7541µV to 16668µV without any decrease in resolution (Figure 3.12) and this was used as the injection volume for the remaining experiments.

3.3.1.6 Temperature

Temperature can have a significant effect on the composition, and therefore chromatography, of organic samples. Sample solutions can potentially degrade during the course of an HPLC analysis, therefore the autosampler temperature can be lowered to slow the rate of degradation. On the other hand, cooling a sample that was prepared at room temperature, particularly those using hot extraction, can cause dissolved or liquid species to solidify in the sample vials, which would lead to changes in sample concentration and could potentially damage the instrument. Temperature also effects the shape, resolution and height of peaks, therefore the column was held at a constant temperature during analysis in order that fluctuations in ambient temperature did not affect reproducibility.

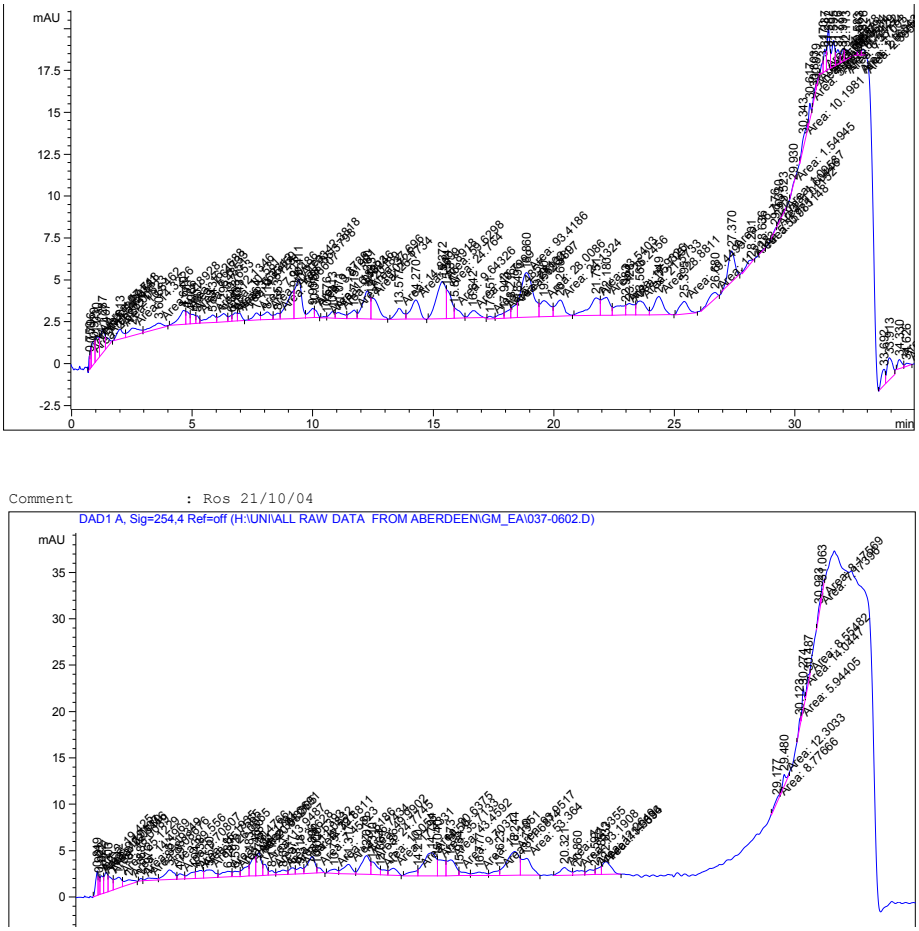
Method

To determine the effect of chilling the autosampler, samples were prepared as per section 3.3.1.1 were injected at room temperature, and then after being equilibrated at 5°C. The methods used were A24 and A26, at ambient and 5°C respectively. Column temperature was initially held at 25°C for HPLC methods S1-S8 (Table 3.1, figure 3.12). Column temperature was then raised to 30°C to ensure constant temperature during the summer months, and this was used as standard throughout the remaining method development using HPLC methods A1-A24, A26 and A27 (Table 3.1). The effect of raising column temperature to 50°C was investigated by preparing a sample according to section 3.3.1.1 followed by analysis using HPLC method A25, with further analysis using HPLC method A26 for comparison.

Results and Discussion

Precipitation of solids occurred when the sample vials were held at 5°C, which did not occur when the autosampler temperature was not controlled, therefore the autosampler temperature was maintained at ambient conditions for the remaining studies. Figure 3.14 shows that increasing the column temperature to 50°C causes the peaks of interest to elute more quickly from the column, with 87% of the peaks eluting before 22min compared to 60% at 30°C, resulting in an unfavourable 13% loss in the average resolution between the peaks obtained at 50°C compared to 30°C, therefore the column was maintained at 30°C in the remaining studies.

Figure 3.14 Chromatography obtained with a column temperature of 30°C (top) and 50°C (bottom)



3.3.2 Sample Preparation

Organic molecules can be highly sensitive to certain environmental conditions, and can decompose or degrade when exposed to factors such as moisture, heat, light or other chemicals. As such, the conditions under which soil is stored and prepared has the potential to affect its organic profile. In order to verify that the analytical method does not introduce

artefacts that could influence the outcome of the comparison, it is also important to monitor the effects of each step in the laboratory procedure on the profiles obtained.

The sample preparation methods detailed in previous studies required 1g of sample, which may not be available in forensic cases and involved several steps, which would take several hours to complete. Complex sample preparation increases cost, reduces efficiency and so reduces the usability of the method for routine use in a forensic laboratory, therefore the sample preparation was redesigned for this study. Equipment performance and chromatography was monitored throughout the method development stage and a series of adjustments to the sample preparation were made accordingly.

Unless stated otherwise, the procedure for preparing HPLC samples was as per section 3.3.1.1 and the sample was injected onto an Agilent 1200 HPLC system with a quaternary pump and PDA detector using the HPLC method A7 parameters detailed in Table 3.1.

3.3.2.1 Drying

Water is a powerful solvent and could lead to the degradation of organic soil constituents through hydrolysis reactions, or by providing a reaction medium to allow alterations to soil composition by other chemical or biological mechanisms. In addition, many soils have excellent moisture retention properties therefore water can account for a substantial proportion of soil's mass and volume. The effect of removing water on the chromatographic profiles was monitored.

Methods

Moisture Content

In order to determine the moisture content of the soil, approximately 25ml of soil was added to a separate pre-weighed glass vial, which was weighed then placed, unsealed to allow moisture egress but protected from light, in an electrically ventilated drying cabinet. The vials were weighed periodically until they were completely dry and had reached a stable weight (stability defined as a change of <0.1 Moisture %w/w between weighings) then the %w/w Moisture was calculated from the loss on drying as defined in Equation 3.1.

Equation 3.1

$$\text{Moisture} = 100 \times \left(\frac{\text{Untreated Sample Weight} - \text{Stable Weight}}{\text{Untreated Sample Weight}}\right) \%\text{w/w}$$

where,

Untreated Sample Weight = Weight of Vial Containing Untreated Sample - Weight of Empty Vial

and

Stable Weight = Weight of Vial Containing Dry Sample - Weight of Empty Vial

Drying Cabinet

One way to remove excess moisture from analytical samples is by placing them in a drying cabinet where they are exposed to gentle heating from incandescent bulbs and increased airflow by way of forced ventilation. This technique could lead to thermal or photo-induced degradation of organic molecules, while the slightly elevated temperature could promote biologically induced changes to the organic composition of the soil. In order to monitor these potential effects, a sub-sample of approximately 5ml well-mixed soil from Locations 1 to 4 were placed in the drying cabinet in open containers until successive weights differed by <0.1% w/w. In addition, a sub-sample from location 4 was kept in a sealed container to prevent loss of moisture and allow for the comparison of chromatograms for moist and dried soil.

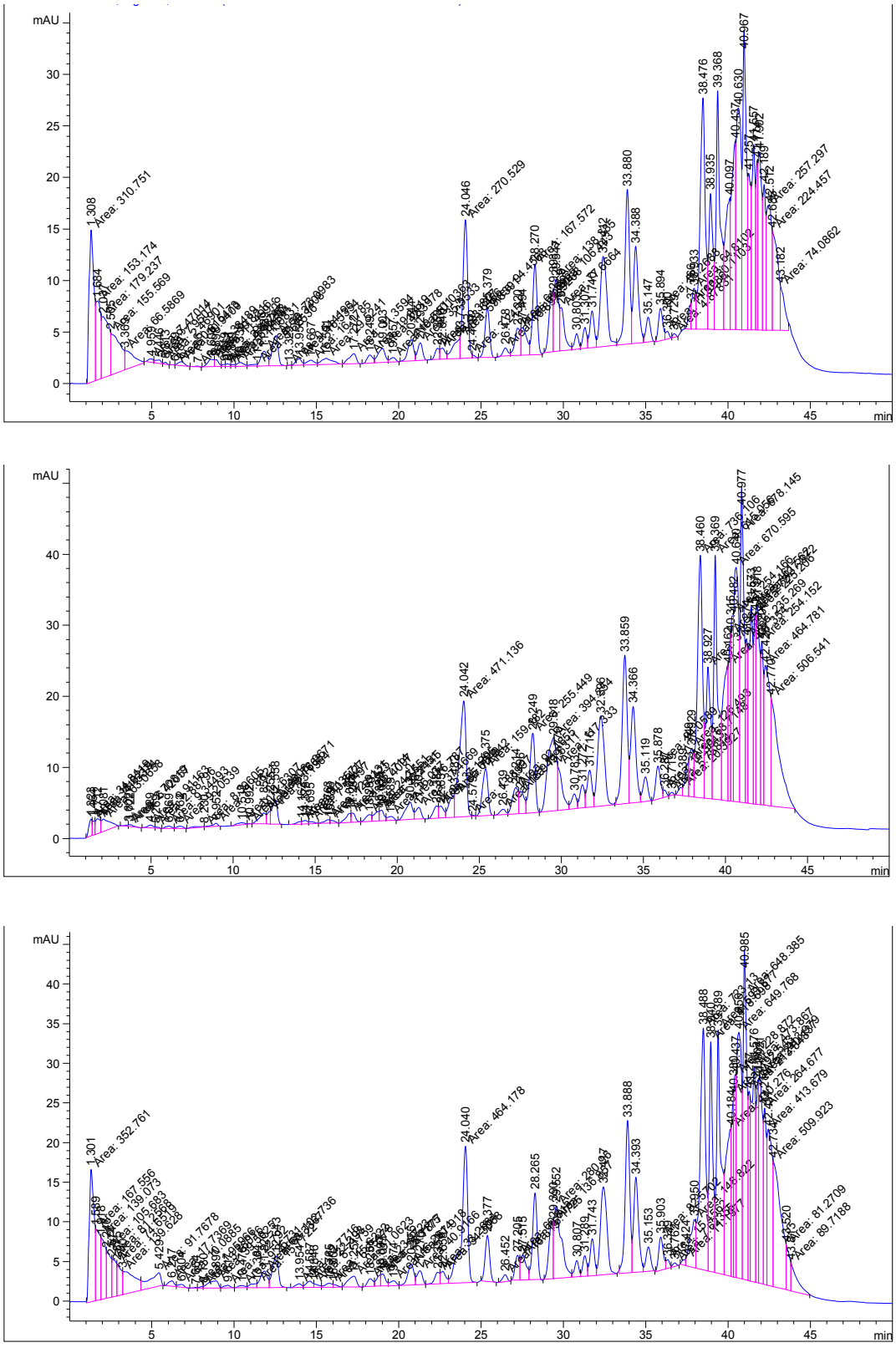
Nitrogen

One way to remove excess moisture from analytical samples is by exposing the sample to a stream of dry, inert gas. This process offers an advantage over drying the samples with heat in that there is no increased risk of thermal degradation, however it can on occasion introduce impurities into the sample, either as a result of the purity of the gas itself, or from the equipment used to deliver the stream of gas, such as rubber tubing. To monitor the effect of drying under Nitrogen gas, soil from Location 4 was mixed well, and a sub-sample was placed in a glass vial and placed under a stream of nitrogen overnight.

Results and Discussion

Locations 1-4 were determined to have moisture contents of 0.19% by weight, 6.35%, 17.92%, 30.95% by weight, respectively and the chromatography of the dried and undried samples are displayed in figure 3.15. There were no significant changes to the chromatograms obtained with either of the drying methods therefore samples were allowed to air dry in the remaining experiments.

Figure 3.15 Chromatograms for samples from location 4 that were not dried (top), dried in the drying cabinet (middle), and dried under Nitrogen (bottom)



3.3.2.2 Working concentration

One of the aims of this project was to reduce the amount of sample required per analysis in order to maximise the potential impact of this technique for use in case work, and it was

therefore necessary to perform experiments to ascertain the effects of reducing the required sample amounts and concentration, as has been discussed by McCulloch et. al. (33).

Method

Bommarito et. al (18) used a working concentration of 1g/ml, therefore approximately 1g was weighed into a centrifuge tube and 1ml acetonitrile was added by automatic pipette and prepared as per section 3.3.1.1. The required sample quantities were then scaled down and HPLC samples were re-prepared using 500mg soil and 500µl acetonitrile (working concentration 1g/ml) to ensure the extraction efficiency was not effected by the reduction in sample amount or solvent volume. The working concentration was then reduced to 500mg/ml, using 250mg soil in 500ul acetonitrile.

Results and Discussion

When samples were prepared at 1g/ml using 1g of sample (Figure 3.12, S2-3), approximately 500ul of supernatant was available for injection onto the HPLC following centrifugation and filtration, which was more than required for analysis. The quantities could therefore be scaled down to 500mg soil and 500ul acetonitrile, as over 200ul of supernatant was available for injection and the chromatograms obtained gave peaks of the same size (Figure 3.12, S4-5) showing that the extraction efficiency was not reduced by this size of reduction in sample amounts or volume of solvent, however this required the use of vial inserts to avoid injecting air bubbles onto the HPLC. Upon increasing the injection volume to 50ul (Figure 3.12, S6-8) it was possible to reduce the sample concentration to 500mg/ml, using 250mg of soil and 500ul acetonitrile, which again resulted in only around 200ul being available for injection onto the HPLC system, however it was not possible to reduce the quantities any further due to unavoidable loss of the sample solution during the filtration step. It may be possible to further reduce the sample quantities in future by preparing more dilute samples and adding in a sample concentration step, however this would add to the complexity of the method, creating scope for analyst error, and would make the sample preparation step more time consuming (and therefore more expensive in a commercial laboratory) and could also require the use of relatively expensive consumables to allow for solid phase extraction, and significant method development to select a suitable internal standard. The working concentration was therefore reduced to 500mg/ml, using 250mg soil in 500ul acetonitrile for the remaining experiments.

3.3.2.3 Extraction Solvents

The solubility of any given compound is dependent on the particular solvent used, and is particularly influenced by the polarity of that solvent, which is in turn determined by the molecular structure. Solvents with polarities similar to the target analyte generally produce

the most efficient dissolution and "like dissolves like" is a useful heuristic for solvent selection. In this case, since the composition of the sample is unknown, a range of solvents suitable for use with Reverse Phase HPLC was tested to determine which extracted the greatest number of compounds and generate the most points for comparison between samples.

Method

An HPLC sample was prepared as per section 3.2.2.1 and analysed by method A7 (table 3.1), a second sample was prepared with 50:50 acetonitrile: water and analysed by method A7, a third sample was prepared with methanol used in place of acetonitrile and analysed by HPLC Method A9, and a fourth sample was prepared with ethanol used in place of acetonitrile and analysed by HPLC Method A11.

Results and Discussion

The effect of the use of methanol, ethanol, acetonitrile and 50:50 acetonitrile: water to prepare the sample extracts is displayed in figure 3.16, which shows the changes to peak shape and resolution resulting from changes to the analytical solvents.

Figure 3.16: Chromatograms obtained for samples extracted in Methanol (Green) Ethanol (Pink) Acetonitrile (Red) and 50:50 Water: Acetonitrile (Blue)

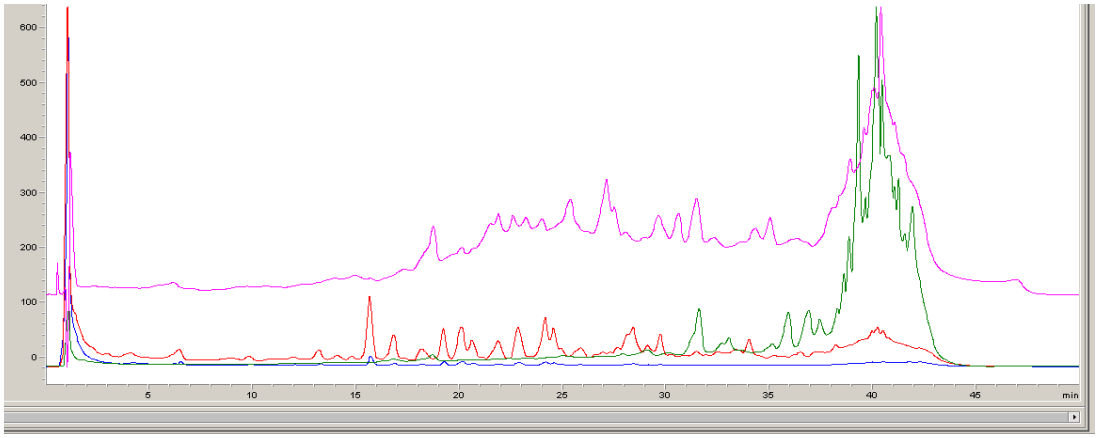


Figure 3.16 shows that acetonitrile was the best extraction solvent as this produced chromatograms with the greatest number of individual peaks, which were the most clearly resolved peaks. The baseline for the acetonitrile samples was stable and therefore produced the strongest signal to noise values, relative to the other solvents used. Acetonitrile was therefore selected as the extraction solvent for the remaining experiments.

3.3.2.4 Extraction:

The ability of the sample preparation method to transfer organic compounds from the sample matrix into solution dictates the amount of sample required to perform the analysis. Extraction efficiency can be improved by maximising the potential for interaction between

solvent and analyte molecules either through mechanical action or increasing the kinetic energy of the molecules through heating.

Methods

Sonication

A common method of extraction involves applying ultrasonic pressure waves to the sample, which produces cavitation in the sample, in which voids are created and oscillate in response to the ultrasonic field, which can in turn disrupt the intermolecular forces preventing analyte molecules from entering the solution phase. In order to assess the effect of sonication on the chromatographic profiles generated, an HPLC sample was prepared as per section 3.3.1.1 and analysed by HPLC method A24 (Table 3.1) for comparison with samples extracted by other means.

Hot Extraction

Thermal energy causes solvent molecules to move more quickly, thus increasing the likelihood of interaction between solvent and solute. Heat can also transfer energy to sample molecules sufficient to overcome the inter and intra-molecular attractions that prevent dissolution, however thermal energy can of course make and break chemical bonds resulting in changes to sample composition. In order to assess the effect of heating on the chromatographic profiles generated, an HPLC sample was prepared as per section 3.3.1.1, heating the samples to 60°C in place of sonication, and analysed by A24 (Table 3.1) for comparison with samples extracted by other means.

Vortex Mixing

Vortex mixing involves oscillating the sample container at such a rate that a vortex is created which forces the solvent through the sample in an orbital motion, mixing the solid and solution phases and thereby increasing the interaction between solvent and solute molecules. The forces created by the vortex can also help to break down aggregated particulate matter to further improve extraction. In order to compare the effect of vortex mixing on the chromatographic profiles generated, an HPLC sample was prepared as per section 3.3.1.1 using the vortex mixer in place of sonication, and analysed by HPLC method A24 (Table 3.1) for comparison with samples extracted by other means.

Grinding

The particulate matter present in soil can create a physical barrier to complete recovery of all soluble compounds present in the sample. In order to maximise the potential for full recovery of soluble organic compounds, it is sometimes appropriate to grind samples prior to

extraction, however this process generates heat and can increase the moisture content of samples, in addition to providing an opportunity for samples to become contaminated in the grinding apparatus, therefore care must be taken that grinding samples does not introduce any changes to the chromatographic profile obtained. In order to test the effects of grinding on the observed chromatography, samples were ground in a pestle and mortar then an HPLC sample was prepared as per section 3.3.1.1 and analysed by HPLC method A24 (Table 3.1)

Results and Discussion

The results of the use of different sample preparation and extraction methods on the chromatography of the samples are presented in figure 3.17.

Figure 3.17: Chromatograms obtained during extraction method development for blank acetonitrile (red), and samples extracted using sonication (green), hot extraction (pink), vortex mixing (dark green), vortex mixing plus hot extraction (blue), grinding (yellow), grinding followed by centrifugation (purple)

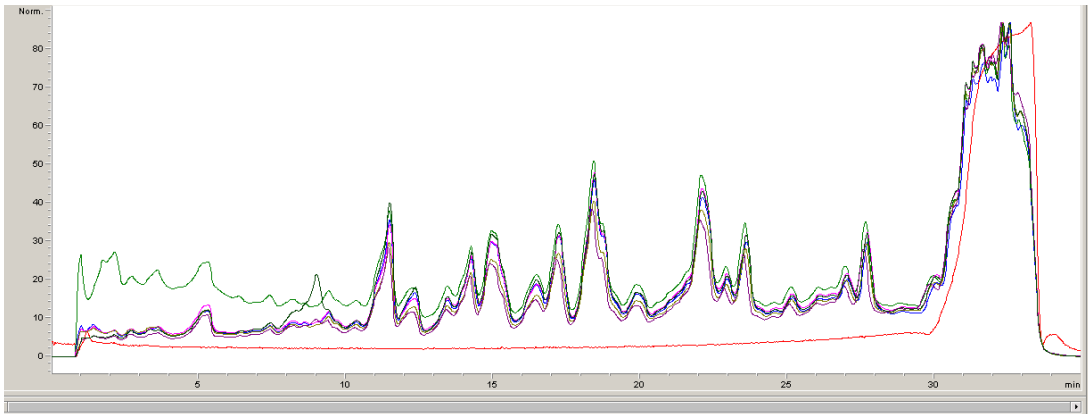


Figure 3.17 shows that there are no significant differences in the peak height, peak shape or the resolution between peaks for the chromatograms obtained with each of the different extraction methods trialled in this study, with the exception of the sample extracted using the vortex mixer, for which a spurious set of peaks was observed at around 9min. Homogenisation of samples during grinding can present problems when interpreting evidence, as discussed in Chapter 2, therefore to avoid these potential issues, and in the absence of clear advantages for any of the extraction methods over the others, and in order to avoid any potential thermal degradation, and the unidentified cause of the spurious peaks in the sample prepared by vortex mixing, sonication was selected as the extraction method for all remaining experiments.

3.3.2.5 Clarification Method

Particulate matter is incredibly harmful to HPLC instrumentation and therefore must be removed from samples to prevent blockages, ensure optimal flow characteristics, avoid

disruption of the UV light path and protect internal components such as the pistons, frits, seals, tubing and flow cell windows. Samples can be clarified by filtration or centrifugation, however both techniques have advantages and disadvantages.

Methods

Centrifugation

Spinning mixed samples at high speed in a centrifuge separates the mixture according to the density of its constituents, which can remove the need for filtration by allowing the liquid supernatant layer to be easily removed, leaving the denser particulates in the sample container, however this may result in a sampling bias in favour of lower density compounds, which may influence the discriminatory power of the method. This removes the opportunity for contamination from filters or adsorption of compounds of interest, and reduces the cost of consumables but adds time to the overall sample preparation, which represents a cost in the form of analyst time. To evaluate the use of the centrifuge to clarify samples, HPLC Samples were prepared as per section 3.3.1.1 from dry, well mixed soil from Location 1, omitting the filtration step, and analysed by HPLC Method A24 (Table 3.1).

Filtration

Filtering the sample through a suitable membrane will remove particulate matter of a given size very quickly, however some dissolved compounds may become adsorbed to certain membrane materials, preventing their detection. Furthermore, compounds present in the filter may be dissolved by the constituents of the sample, for instance the solvent, and pass through the membrane and into the sample, creating artefacts in the chromatogram. To verify the absence of artefacts, and that compounds of potential interest are not retained by the filter, the chromatography obtained from filtered samples, prepare as per section 3.3.1.1 and analysed by HPLC Method A24, was compared with that obtained for unfiltered samples used to verify the effectiveness of centrifugation.

Results and Discussion

Figure 3.18 shows that centrifugation of the samples resulted in visibly clear and particulate free sample solutions, suggesting that centrifugation was adequate to remove particulate matter.

Figure 3.18: Samples Prior to Centrifugation (top) Samples after centrifugation (bottom)

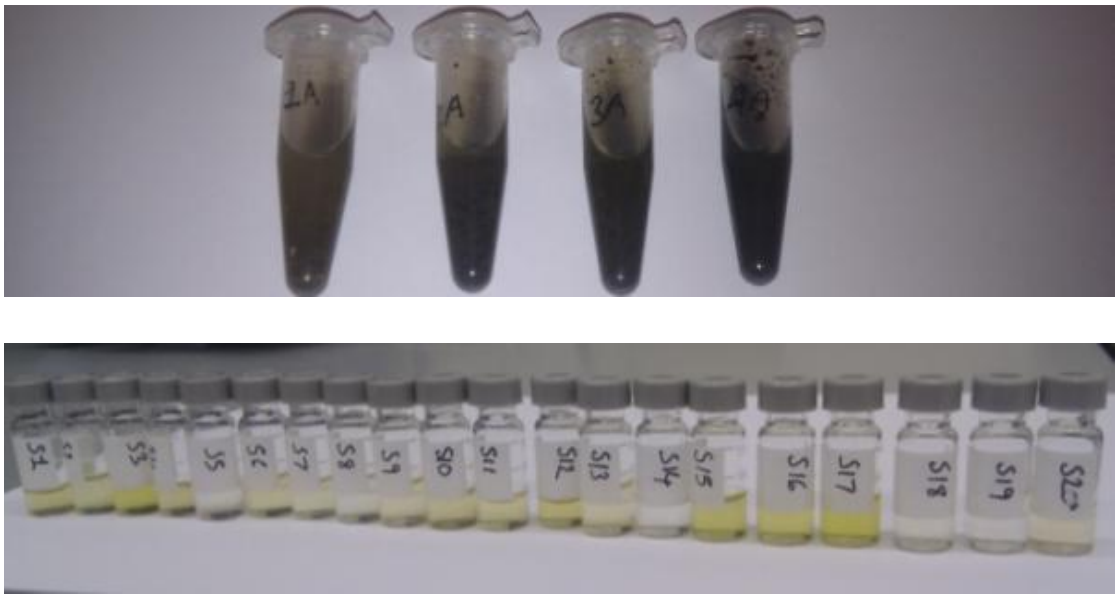
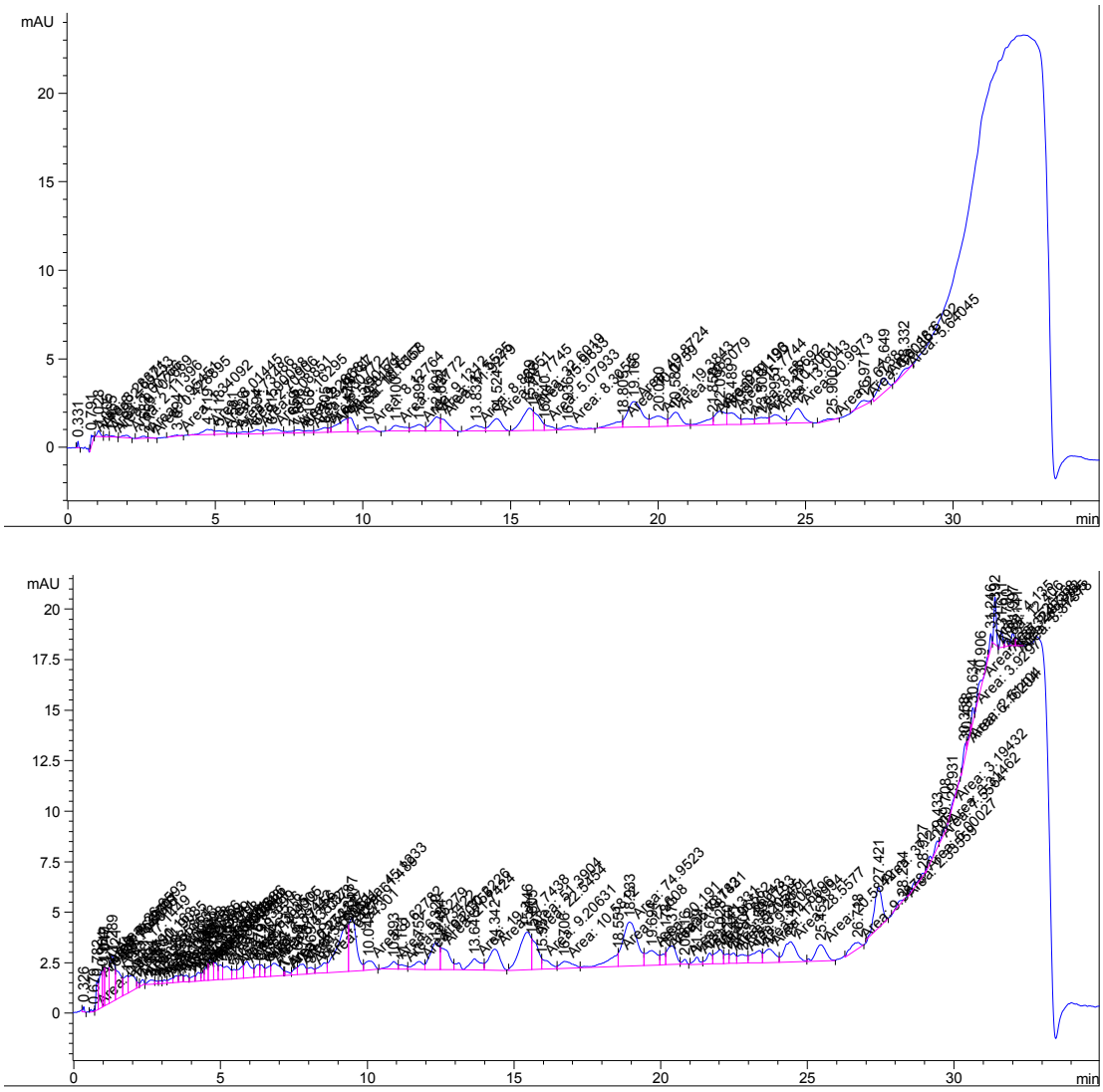


Figure 3.19 shows there were no visible differences in the chromatography obtained using the filters, indicating that there was no leaching of chemicals into the sample from the filter and that the compounds of interest were not adsorbed by the filter.

Figure 3.19 Chromatography obtained for samples centrifuged only (top) and for filtered and centrifuged samples (bottom)



There were no differences observed in the chromatograms obtained for filtered and unfiltered samples, and the sample solutions were free from visible particulates, therefore samples were not filtered for the remaining method development experiments, however an increase in system backpressure was observed, indicating that particulate matter was accumulating in the system after multiple injections, therefore the filtration step was reintroduced to the method and no further pressure issues were encountered.

3.3.3 Sample Storage

It is essential to the correct interpretation of forensic evidence that the effect of the passage of time on the outcome of forensic analyses is monitored, since certain types of sample may become degraded beyond recognition, rendering meaningful comparisons useless, under certain combinations of environmental conditions, for instance temperature, humidity and packaging materials. There can be no control over the length of time taken to detect and fully

investigate a crime, and cold case reviews have proven to be highly successful in solving complex cases when new techniques are developed or new evidence is discovered, for instance in miscarriages of justice, appeal cases or unsolved crimes. It may, therefore, be necessary to store forensic evidence for decades in order to pursue justice to its limits, for instance in complex cases and serious crimes. Storage conditions also directly influence the cost of retaining evidence, and refrigeration or freezing of samples increases the cost of storage considerably, therefore the effects of storing the samples under different storage conditions must be investigated

3.3.3.1 Solution Stability

In a regulated laboratory, it is sometimes necessary to re-inject and re-vial samples if discrepancies or abnormalities are observed in the data, and since it may not be possible to perform data processing and peer review immediately, several days may have elapsed between initial sample preparation and repeat analysis. Determining the length of time a prepared sample can be stored without significant changes to the chromatography is a common stage in the method development and validation of organic analyses (152) and becomes even more pertinent in forensic analyses where it is not always possible to freshly re-prepare a sample due to limited sample quantities. In order to assess the stability of sample solutions stored under different conditions the following analyses were performed:

Methods

Fridge

Samples from points 1A, 2A, 3A and 4A, which were originally prepared using the method described in section 3.3.1.1 and analysed by HPLC Method S8 (Table 3.1), were stored at 2-8°C for seven days then passed through a 0.22µm syringe filter into an HPLC vial containing a 200µl insert and re-injected using the same HPLC parameters.

Freezer

Samples were prepared and analysed as described in section 3.3.1. 1 After injection, the vials were stored at -20°C for 14 days then re-injected using the same HPLC parameters (method A24).

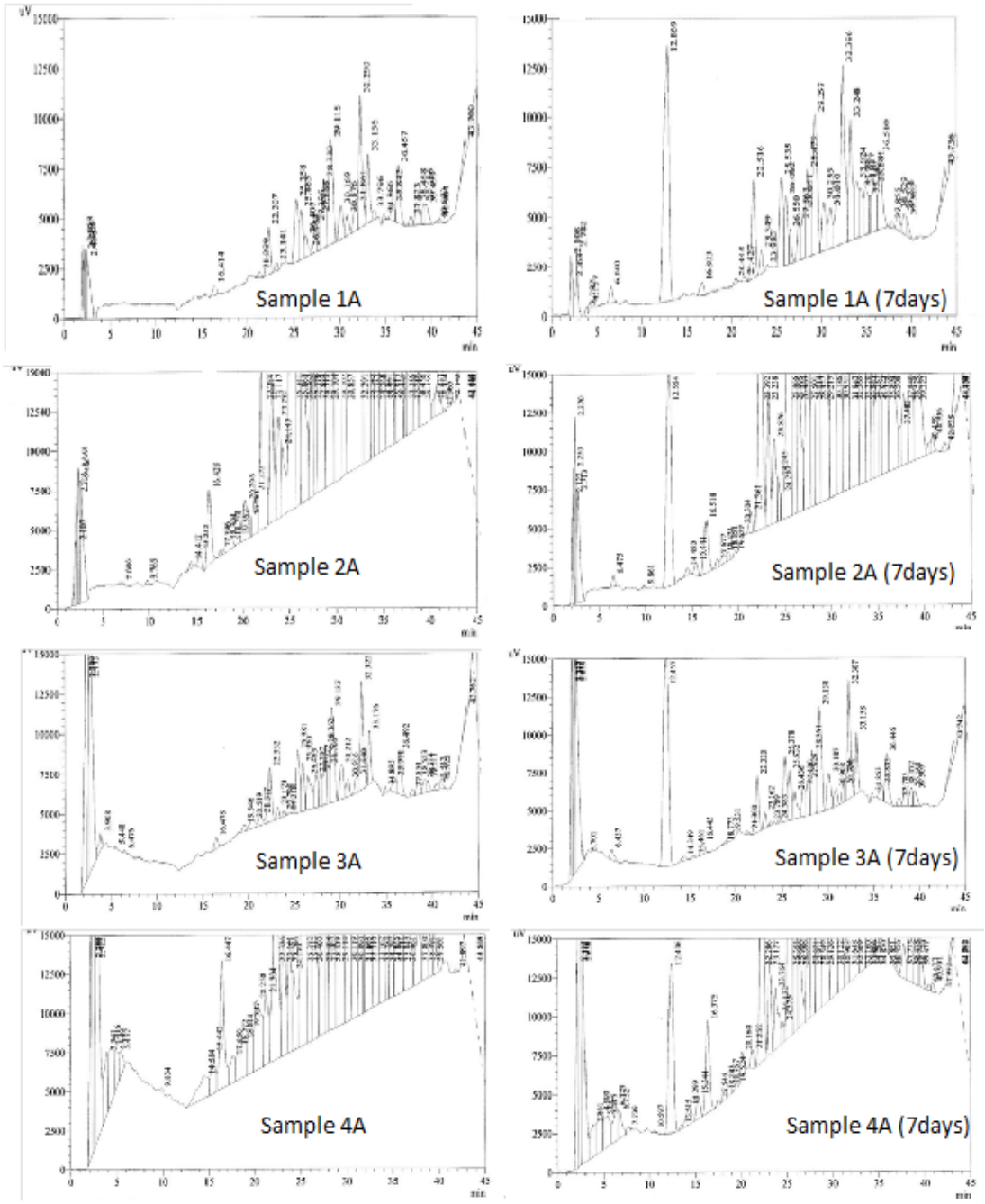
Results and Discussion

The results obtained after 7 days refrigeration are shown in figure 3.20 and summarised in table 3.2, while the results for frozen samples are displayed in figure 3.21

Fridge

As discussed by McCulloch et al. (33) and shown in figure 3.20, solutions were not stable in solution when refrigerated for 7 days. Chromatograms of sample solutions from each location showed qualitative changes upon storage, most notably through the appearance of additional peaks with retention times around 6 min and 12 min, and changes in relative peak sizes.

Figure 3.20: Solution stability for samples from each sample location after 7 days refrigeration



Quantitative changes were observed in sample solutions prepared from all locations after 7 days refrigeration. Evidence of the formation of new compounds was observed through the

presence of additional peaks in all solutions upon reanalysis after 7 days. The changes in the peak areas detected in each sample indicate that the compositions of the samples are substantially altered in solution and that these effects are not consistent across all sample locations. Newly formed compounds accounted for 18.8%, 3.9%, 16.9% and 4.5% of the total sample for samples 1A, 2A, 3A and 4A, respectively. The quantity of the compound eluting at around 16 min did not change significantly with respect to the total chromatogram area, however the relative change compared to the initial peak size was, again, substantial and inconsistent across the sample locations.

Table 3.2 Quantitative differences for samples after 7 days refrigeration

	Sample	Total Peaks	Total Area (µV.s)	Area (µV.s) RT ~ 6min	Area (µV.s) RT ~ 12min	Area (µV.s) RT ~ 16min
Initial	1A	32	1381333	N/D	N/D	14160 (1.9%)
7 days		35	2838697	20098 (0.7%)	51470 (18.1%)	18636 (0.7%)
%Change vs. Initial		9.4	105.5	N/A	N/A	24.0
Initial	2A	48	14788856	N/D	N/D	144330 (1.0%)
7 days		49	14277982	15500 (0.1%)	537414 (3.8%)	105653 (0.7%)
%Change vs. Initial		2.1	-3.5	N/A	N/A	-26.8
Initial	3A	37	2744890	N/D	N/D	17551 (0.6%)
7 days		38	2990580	19288 (0.6%)	486582 (16.3%)	21340 (0.7%)
%Change vs. Initial		2.7	9.0	N/A	N/A	21.6
Initial	4A	47	14930434	N/D	N/D	285285 (1.9%)
7 days		54	10005195	79634 (0.8%)	374060 (3.7%)	176291 (1.8%)
%Change vs. Initial		14.9	-33.0	N/A	N/A	-38.2

Values in parentheses indicate the contribution of the peak to the total chromatogram area
N/D= Not Detectable, N/A= Not Applicable.

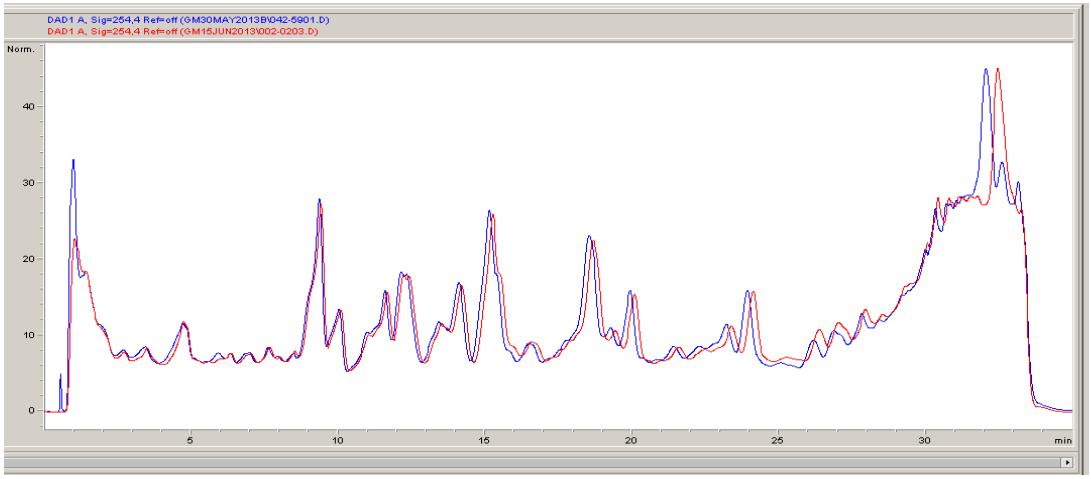
All refrigerated samples exhibited both qualitative (Figure 3.20) and quantitative (Table 3.2) changes after 7 days storage in solution at 2-8˚C. New peaks formed in all solutions, most notably the large peaks at around 6min and 12min. Quantitative changes were observed in the peak eluting at around 16 min. The magnitude and direction of this change varied

between samples, indicating that each sample exhibited different solution chemistry leading to different rates of formation and degradation of this peak.

Freezer

The profile of the frozen sample solution (Figure 3.21) was almost completely unchanged after two weeks with the exception of a shift in the retention time of a large peak at the end of the chromatogram, possibly masking two smaller peaks. These results suggest that the solutions are unstable under refrigeration but are largely stable for two weeks when stored at -20˚C, however further investigation of sample stability is recommended for future studies.

Figure 3.21: Chromatograms obtained upon initial analysis (Blue) and after 14 days storage in solution at -20˚C (Red)



3.3.3.2 Storage Conditions

Heat is a key factor in reaction kinetics, hence many organic chemicals are vulnerable to heat exposure, through evaporation or degradation by thermally mediated degradation mechanisms. In order to compare the effect of storage at room on the chromatographic profiles generated, HPLC samples were prepared from soil stored under different conditions.

Methods

Ambient

Approximately 5g well-mixed soil was added to a brown paper envelope and stored under ambient conditions, in a sealed glass container but protected from light. Samples were analysed by HPLC method A24 after 14 days

Light Exposure

Approximately 5g well-mixed soil was added to a clear glass vial, and stored under ambient conditions, exposed to light. Samples were analysed by HPLC method A24 after 14 days.

Fridge

Approximately 5g well-mixed soil was added to an LDPE bag and stored at approximately 5˚C, protected from light. Samples were analysed by HPLC method A24 after 14 days.

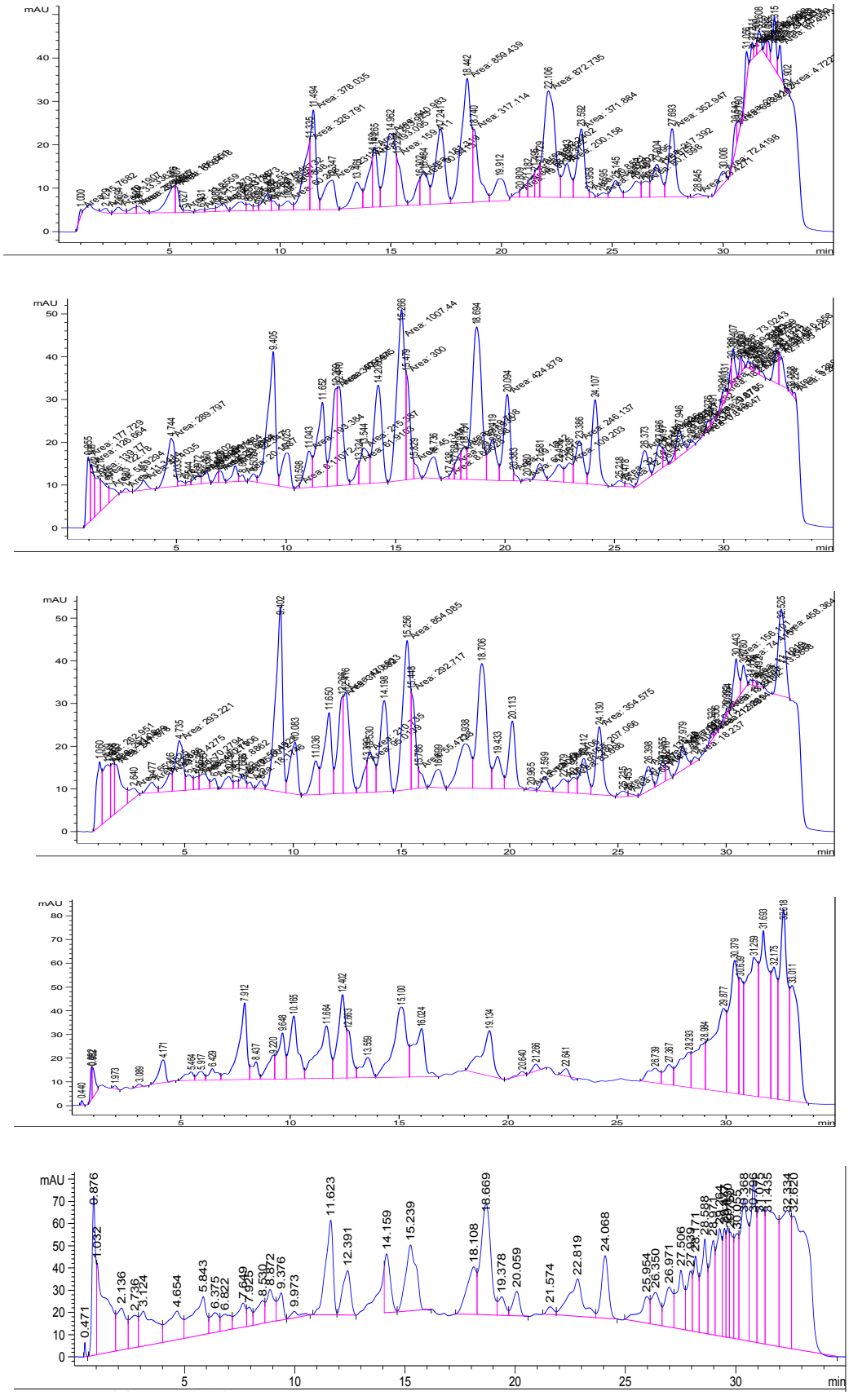
Freezer

Approximately 5g well-mixed soil was added to an LDPE bag and stored at approximately -20˚C, protected from light. Samples were analysed by HPLC method A24 after 14 days.

Results and Discussion

Figure 3.22 shows that each of the different storage conditions resulted in changes to the chromatography, and each had different effects on the profiles therefore further investigation of sample stability and optimum sample storage conditions is required. Care was taken to store all samples consistently and analyse samples as soon as practically possible (no more than 72 hours) after collection in all remaining experiments. In addition, artefacts with a mould-like appearance were observed to have contaminated the samples that were stored in sealed glass containers, and erucamide, a plasticiser used in packaging materials, was observed in the mass spectrum of soils stored in LDPE bags, which could cause interferences in any future attempts to identify samples by LC-MS therefore samples were stored in brown paper for all remaining experiments.

Figure 3.22 Chromatograms obtained (from top to bottom) for a freshly prepared sample, a frozen sample, a refrigerated sample, a sample stored at ambient conditions protected from light, a sample stored at ambient conditions



3.4 Conclusions

It was not possible to obtain chromatographic separation using the HPLC parameters defined by Bommarito et. al. (18) therefore all HPLC parameters were redefined to provide optimum chromatography. C18 column stationary phase provided the best chromatogrphic resolution, while the use of gradient elution using acetonitrile and water allowed the analytical run time to be reduced by 65% and it was determined that no improvement to peak shape or resolution was achieved with the addition of acid to the mobile phase. It was possible to improve the sensitivity of the method by increasing the injection volume to 50 µl, while it was determined that maintenance of the column temperature at 30°C provided preferable resolution to heating the column. The sample preparation technique was simplified as the method development experiments showed there to be no benefit from drying the samples with nitrogen or heat, and that the best method of extraction was sonication using acetonitrile as the extraction solvent. Centrifuging the samples prior to a single filtration step was shown to remove particulate matter effectively and prevent the problematic rise in instrument pressure noted by Bommarito et. al (18) which negated the requirement for a guard column and additional syringe filters. The sample composition was shown to be unstable upon storage in solution, and sealed transluscent containers such as glass vials or LDPE bags, were noted to cause the growth of mould-like artefacts therefore breathable, opaque containers were selected for use in the final method.

On the basis of the conclusions of the method development experiments, two different sets of analytical methodology were selected for use in the experiments detailed following chapters, which are specifiied in sections 3.4.1 and 3.4.2. The first, preliminary method was used to assess the feasibility of the use of HPLC and for comparison of the use of HPLC with QGSTA (Chapters 4 and 7) and the final method was the result of the later stages of method development and was used for all other experiments (Chapters 5, 6 and 7) and these methods are summarised in sections 3.4.1 and 3.4.2.

3.4.1 Preliminary Method

Samples were stored in LDPE bags at -20°C which were removed from storage and allowed to equilibrate to room temperature prior to analysis. Approximately 250mg soil was weighed into a polypropylene tube and 500µl Acetonitrile was added by automatic pipette. The samples were then mixed manually and sonicated for 20 min at ambient temperature. The samples were centrifuged at 13,000 rpm for 15 min and the supernatant sample solution passed through a 0.22µm PTFE syringe filter into an HPLC vial containing a 200µl insert for analysis.

50 µl of each sample was injected onto a Whatman Partasil 10 ODS analytical HPLC column (10um particle size, C18 packing material, column dimensions 250 x 4.6mm) held at 25°C, with a DAD collecting UV-Visible spectra from 190-800nm, monitoring at 254nm and sample vials were held at 10°C during analysis. Samples were eluted at a flow rate of 1ml/min with water and acetonitrile using the gradient detailed in table 3.3

Table 3.3: HPLC Gradient for Initial Analysis

Time(minutes)	% Water	%Acetonitrile
0	45	55
5	45	55
35	75	25
38	98	2
40	98	2
45	45	55
50	45	55

This preliminary method was used for the Feasibility study (Chapter 4) and comparison with QGSTA (Chapter 7) and by McCulloch et. al. (33)

3.4.2 Final Method

Samples were stored in brown paper envelopes at ambient conditions and allowed to air dry prior to analysis. Within 72 hours of sampling, approximately 250mg soil was weighed into a polypropylene tube and 500µl acetonitrile was added by automatic pipette. The samples were then mixed manually and sonicated for 20 min at ambient temperature. The samples were centrifuged at 13,000 rpm for 15 min and the supernatant sample solution passed through a 0.22µm PTFE syringe filter into an HPLC vial containing a 200µl insert for analysis.

50 µl of each sample was injected onto a Waters Xbridge HPLC column (3.5um particle size, C18 packing material, column dimensions 100 x 4.6mm) held at 30°C, with a DAD collecting UV-Visible spectra from 190-800nm, monitoring at 254nm and sample vials were held at 10°C during analysis. Samples were eluted at a flow rate of 1ml/min with water and acetonitrile using the following gradient in table 3.4:

Table 3.4: HPLC Gradient for Final Analysis

Time(minutes)	% Water	%Acetonitrile
0.0	53	47
3.0	45	55
24.0	26	74
29.0	2	98
31.0	2	98
32.0	53	47
35.0	53	47

This final method was used for the selection of marker peak sets (Chapter 5), the investigation of seasonal and geographical variability (Chapter 6), and the comparison with wax marker profiling by GC (chapter 7) (see also McCulloch et. al. (35)).

4 Feasibility Study

4.1 Introduction

The purpose of this study was to evaluate the performance of a new HPLC technique following redevelopment of some of the sample collection, preparation, analysis and interpretative approaches, as described in Chapter 3, in order to assess the feasibility of the use of HPLC for soil analysis in a more forensically relevant context. The re-developed method was intended to be appropriate for comparing trace soil samples for the purposes of excluding crime scene, alibi site and unknown samples. These contextual details are important, since the priorities in civil or environmental forensic cases, those involving bulk samples and those which aim to predict the geographic provenance of a sample are significantly different, as are the considerations required for correct interpretation of the evidence (107,14,16). This study has been presented by McCulloch et. al (33) and was designed to address whether it is possible to distinguish groups of trace soil samples obtained from locations that are located in close proximity to one another, and thus forensically relevant to a crime reconstruction, using a practical, user friendly HPLC analysis.

4.2 Methodology

Samples were collected from Brockwell Park, London, as described in section 3.2. Each sample was assigned an anonymous identification number by a second analyst in order to minimise cognitive bias of the primary analyst during sample preparation and data analysis, the log cross referencing identification numbers and their sample position was kept confidential throughout sample preparation and initial data analysis. Since organic compounds can be thermally labile and soil temperature is known to influence biological activity (4) (169), samples were stored in LDPE bags at -20°C in order to prevent changes to the organic composition of the samples caused by micro-organisms or thermal degradation. Anonymised samples were removed from storage and allowed to equilibrate to room temperature prior to analysis.

This experiment was performed at an early stage in the project, to assess the feasibility of HPLC profiling, therefore the method development experiments described in Chapter 3 were only partially complete. As such, the methodology used in this experiment was an early version of the method used in later experiments as outlined in chapter 3 (section 3.4.1), and is summarised below for clarity

Approximately 250mg soil was weighed into a polypropylene tube and 500µl Acetonitrile was added by automatic pipette. The samples were then mixed manually and sonicated for 20 min at ambient temperature to extract the organic compounds from the soil matrix. In order to remove suspended particulates, the samples were centrifuged at 13,000 rpm for 15 min and the supernatant sample solution passed through a 0.22µm PTFE syringe filter into an HPLC vial containing a 200µl insert for analysis.

50 µl of each sample was injected onto a Shimadzu VP series HPLC system, which was comprised of an SIL-10AD autosampler with cooling tray, a FCV-10AL solvent mixing system, a DGU-14A vacuum degasser, a SCL-10A control unit, a SPD-M10A diode array detector (DAD) and a CTO-10AS column oven. The column used was a Whatman Partasil 10 ODS analytical column (10um particle size, C18 packing material, column dimensions 250 x 4.6mm) held at 25°C in order that fluctuations in ambient temperature did not affect reproducibility. The DAD collected UV-Vis spectra from 190-800nm with monitoring at 254nm and sample vials were held at 10°C during analysis to minimise sample degradation during analysis. Gradient elution was used in order to speed up elution of strongly retained compounds and improve the practicality of the method, the gradient is detailed in Table 4.1, below. UHQ water and HPLC Gradient grade Acetonitrile were used and equipment was subject to a Good Clinical Laboratory Practice maintenance schedule.

Table 4.1: HPLC Gradient for Feasibility Analysis

Time (minutes)	% Water	% Acetonitrile
0	45	55
5	45	55
35	75	25
38	98	2
40	98	2
45	45	55
50	45	55

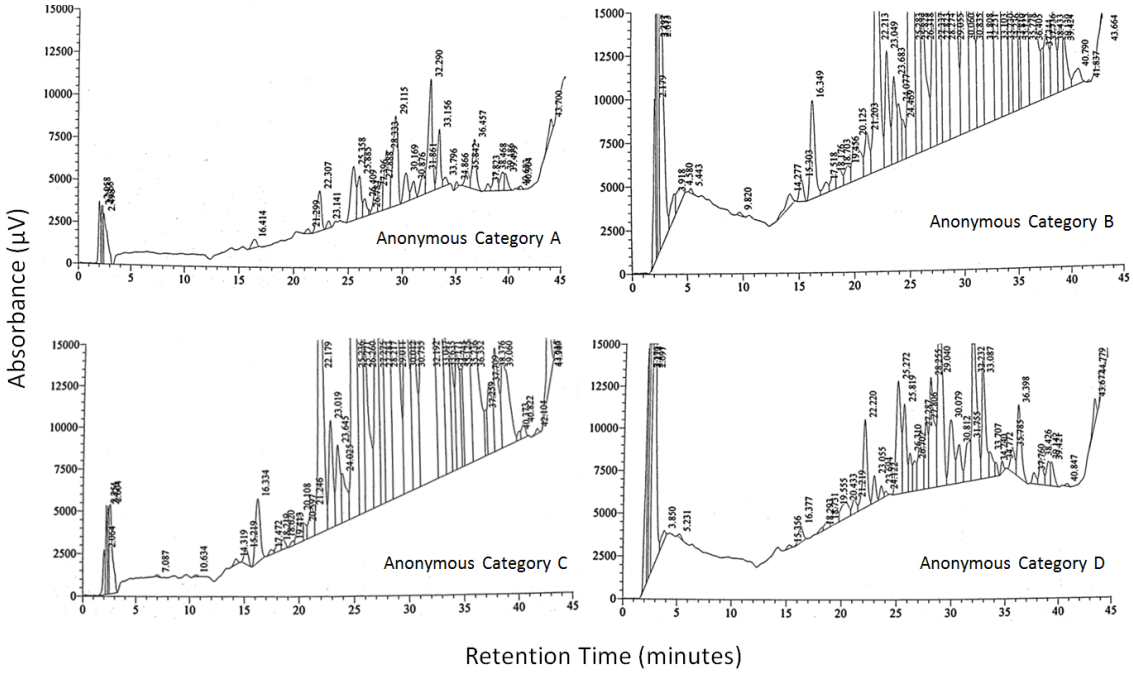
Samples were injected in a random order based on their identification number in order to prevent any systematic instrumental bias.

Data processing was performed by the Shimadzu LC Solution Data Analysis software package and peaks were detected and identified on the basis of predetermined acceptance criteria to ensure a consistent approach was used for all samples. The anonymised sample chromatograms were then subject to visual assessment of the similarities and differences between samples, in addition to Canonical Discriminant Function Analysis (CDFA) in SPSS. This is a multivariate statistical analysis tool which uses linear combinations of the values for each of the variables (chromatographic peaks) to maximise the differences between groups, then provides scatter plots to enable the investigator to visualise the degree of difference between the samples based on their scores for each function, and also generates data on the statistical significance of the sample groupings, and the accuracy with which samples could be assigned to groups based on the canonical functions. Further discussion of the data analysis approaches used throughout the project is provided in Chapter 5, however for this feasibility study, each individual chromatographic peak observed in the entire data set was used as a variable in the data analysis.

4.3 Results

Four distinctive chromatographic profiles were observed upon systematic examination of the chromatograms of anonymised samples, and examples of each type of profile are displayed in figure 4.1.

Figure 4.1 Visual comparison of chromatograms



Combinations of characteristic baseline features and groups of peaks were common to samples taken from the same location, which allowed samples to be grouped visually with 100% accuracy, and the features used for classification are shown in table 4.2.

Table 4.2: Classification of samples by visual comparison of chromatograms

Anonymous Category	Anonymised Samples Present	Sample Points (respectively)	Major Peaks Present		
			3-5min	17-18min	20-21min
A	5, 8, 14, 18, 19	1E, 1B, 1D, 1C, 1A	No	No	No
B	1, 2, 4, 9, 10	4E, 4B, 4D, 4A, 4C	Yes	Yes	Yes
C	6, 7, 9, 13, 20	2D, 2A, 2C, 2B, 2E	No	Yes	Yes
D	3, 11, 15, 16, 17	3C, 3E, 3A, 3B, 3D	Yes	No	Yes

The results of the CDFA for the HPLC data determined three functions with which to categorise the samples. The scores calculated for each sample for the first two canonical functions are shown in the scatter plots in figure 4.2

Figure 4.2 Canonical discriminant function analysis

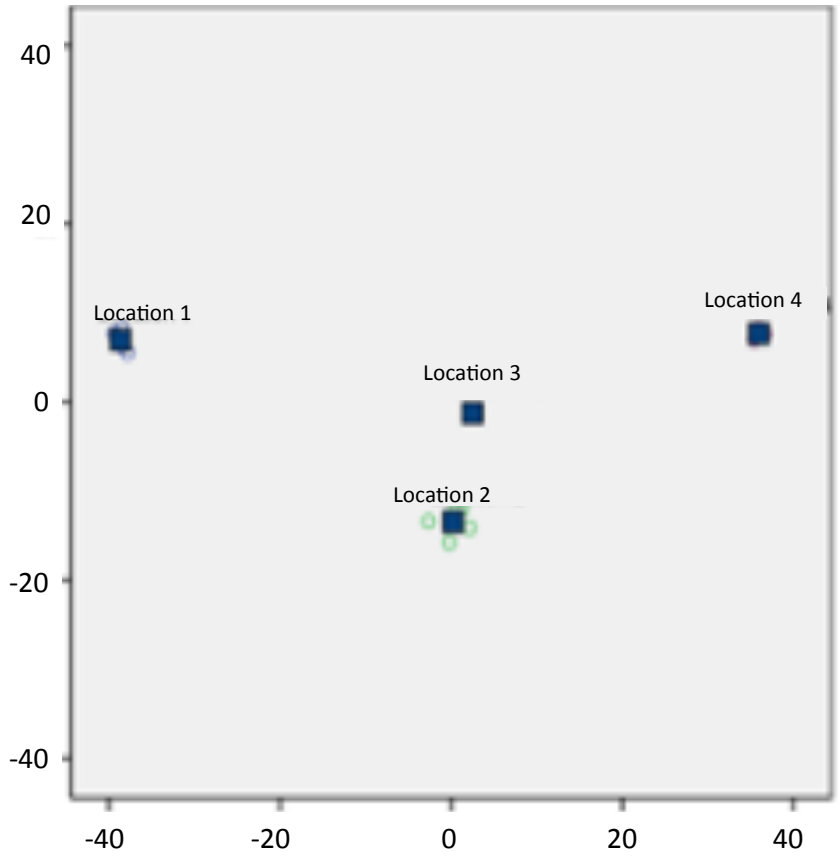


Figure 4.2 Scatter Plots showing the relative positioning of each sample on the basis of the first (x-axis) and second (y-axis) canonical functions generated from the data. The scores for the group centroids are indicated by a blue square.

Functions 1, 2 and 3 explained 89.5%, 9.3%, and 1.2% of the variance between the sample groups, respectively and were significant at the 99% confidence level (p=0.000). Again, 100% of samples were classified correctly with these functions.

4.4 Discussion

4.4.1 Visual Comparison

Systematic visual inspection of the samples' chromatographic profiles resulted in the correct grouping of all samples (Figure 4.1). The criteria used to assign samples to groups are detailed in Table 4.2. All samples at Location 1 could be easily distinguished from the remaining samples. Samples from Location 3 were more time consuming to compare, nevertheless

assignment was made with relative confidence. The number of peaks in the samples made visual comparison more difficult for Locations 2 and 4. This process was, naturally, highly subjective; the definition of "major peaks" (Table 4.2) was informed but technically arbitrary and it was noted that, within Locations 2, 3, and 4, some internal variability in terms of relative peak height and shape were discernible. This subjective nature of this approach means that it was not reasonable to apply any statistical treatments to the results in order to assess the significance of the differences between the groups with respect to intra-location variability. It is, however, encouraging that classification of such complex data to such a high degree of accuracy was possible, particularly in a blind trial. As with other techniques, such as microscopy based analyses, which require detailed visual examination by an analyst, the classification process was very time consuming, which would need to be taken into account when considering the practicality of this approach for data analysis in routine forensic analyses. In this study it took around two hours to compare the chromatograms, which somewhat limits the practicality of this approach to data analysis for routine forensic analyses.

4.4.2 CDFA

In order to enable quantitative assessment of the significance of the differences between the sample groups, data were analysed using Statistical Package for the Social Sciences (SPSS) version 17.0. The height for each peak was used to classify samples using CDFA, which has been reported previously in geo-forensic literature for the analysis of similar geochemical data (107). CDFA analysis correctly grouped the samples by location (100%) for all three variables. The results of this analysis are displayed in Figure 4.2 and the results show that the discrimination identified between the four locations was statistically significant at the 99% confidence level for each variable.

The CDFA plot (Figure 4.2) shows that there is greater inter-location variability in comparison to the intra-location variability using all the peaks as variables in the analysis, however some locations were distinguished to greater degrees than others. Location 2 was distinct from all other locations in having a greater degree of between-samples variability, and therefore was not as clearly excluded from the other locations, and Locations 2 and 3 were much more similar to one another than the remaining groups. The greatest difference was between Locations 1 and 4, which were approximately equidistant to Location 2, and Locations 3 and 4 were marginally more similar than Locations 1 and 3.

The most clear differences between the sites that could explain these differences were the nature and extent of vegetative cover (Chapter 3 Figures 3.2, 3.3, 3.4, 3.5), the exposure of the

site to sunlight and the pH and moisture content of the soils, however further investigation would be required to determine the exact sources of the similarities and differences observed.

4.5 Conclusions

The results of this study are extremely promising due to the high level of accuracy with which samples were discriminated and the improvements to the methodology. For the first time, it has been demonstrated that it is possible to discriminate between soil samples taken from different locations within the same broad geographic site by HPLC with high levels of accuracy, which are essential to the evidential value of a technique for forensic applications. These results represent significant improvements in both the accuracy and spatial precision presented in previous studies involving the use of HPLC for the comparison of forensic soil samples; the spatial variability in the locations chosen for this study was considered to be more representative of geoforensic casework and none of the previous studies documented in the published literature have been able to successfully differentiate between all sample locations. The findings from this study indicate that it is possible that, with careful interpretation of results, HPLC could indeed be used in crime reconstruction at the level of excluding specific locations within a crime scene, for instance entrance and exit points or the route travelled between points..

Furthermore, the practicability of the HPLC method has been improved greatly by successfully simplifying the sample preparation, reducing the sample preparation time by a minimum of two hours per sample and reducing the run time by 50%, compared to the most recently reported study in the literature (18). The benefits of these methodological improvements should not be underestimated as the resulting financial savings and increased productivity increases the potential impact of this research in the wider forensic community. In light of the transition to the provision of forensic services by commercial organisations in the UK, techniques that require equipment or skills that are not available in standard analytical laboratories, or are particularly costly, will be of little use in routine analyses. Likewise, high analysis costs limit the extent of research and development that can be performed in an academic environment.

While visual assessment of anonymised samples was shown to correctly assign samples to their location group for all of the samples in this study, it will be necessary to formalise the process of comparing chromatograms to prevent subjective interpretation of the results and address the limitations of visual comparison discussed in section 4.4.1. Further exploratory data analysis is required in order to select the most useful and discriminatory chromatographic

features in order to remove the need to classify the samples by visual examination of the chromatograms. In particular, identification of the peaks that are most consistent within groups and most different between groups would be beneficial in order to reduce the number of peaks used in the comparison of samples, which would in turn make any visual assessment easier and less time consuming, and would also improve the speed at which sample data could be prepared for CDFA.

Further empirical studies will be required in order to ascertain the spatial ranges within which soil chromatographic profiles exhibit homogeneity and define the exact precision limits for successful discrimination of soil locations by HPLC. In order for correct comparisons to be made between samples of known and unknown provenance, it will also be necessary to examine how the chromatography changes over time, in response to various environmental conditions and how well it is preserved in and recovered from various different substrates or matrices. It will also be necessary to demonstrate the ability to successfully discriminate locations within a range of different sites to ensure the validity of the use of the technique in a wider variety of crime scenarios. Nevertheless, the results of this preliminary study indicate that HPLC analysis is indeed a feasible and effective way of accurately comparing forensic soil samples.

It was demonstrated that HPLC is applicable to the analysis of soil samples that are taken from forensically relevant sample points, with respect to both the precision in discriminating geologically similar locations within a small spatial scale and the amount of sample required. HPLC has, therefore, been shown to offer enormous potential to add to the suite of geoforensic techniques currently used to assist the investigation and detection of crimes, and is particularly beneficial since it expands the range of organic analyses available, which provides an independent form of analysis to complement other methods of physical analysis of soil or sediment samples.

5. Data Analysis Development

5.1 Introduction

The research outlined in chapter 3 has addressed the practical problems of previous research, by reducing the sample amount and simplifying the analytical procedure, while the research into the feasibility of using HPLC for geoforensic analysis discussed in chapter 4 demonstrated that the new technique offers excellent discrimination at a spatial scale of relevance in a forensic context, on soils, positioned at close-proximity locations for instance the entrance and exit locations of a crime scene and a possible alibi sites. The newly developed method was found to be appropriate for comparing trace soil samples for the purposes of excluding crime scene, alibi site and unknown samples in criminal cases and these contextual details are important, since they affect the considerations required for correct interpretation of the evidence. Interpretation of forensic evidence is a pertinent issue that remains to be fully addressed by forensic science researchers, despite significant improvements in technical capabilities resulting in ever more complex data sets. There is still a significant need for published research that assists investigators in correctly analysing their data to draw the correct conclusions from their results, with an appropriate consideration of the statistical significance of any evidence presented in court, since "physical evidence cannot be wrong; it cannot perjure itself…..only in its interpretation can there be error" (Kirk, 1974: p4) (40)

Due to the complexity of the structure and composition of geoforensic materials, which provides multiple criteria for comparison, there are a number of different ways in which geoforensic trace evidence can be described and classified (7,8), and it is essential that the complex data sets obtained from testing geoforensic evidence are analysed and interpreted correctly. In the feasibility study outlined in Chapter 4, a great many analytes were detected in the soils tested, which poses a significant barrier to implementation of the technique outside of academia as the required data analysis was too labour-intensive to be considered practical for routine analyses or large numbers of samples. The aim of the work described in this chapter was to select a significantly reduced number of useful target analytes for multivariate statistical analysis, in order to provide the same level of discrimination between sites offered by the existing method, but with a much more easily practicable data analysis method, allowing the use of HPLC as a profiling tool for geoforensic samples in routine casework.

5.2 Methodology

All sites were well-established municipal parkland, intended and maintained for public recreational use by the local authority. Three sites in the UK were selected, Brockwell Park in London, Lochend Park in Edinburgh, Craigiebuckler Estate in Aberdeen, in addition to one site in the USA, Central Park in New York City. The three additional sites were chosen for comparability with the Brockwell Park site in London used in the method development and feasibility study, and each of the additional sites contained areas of similar land use to the four locations used in the previous chapters and described in table 5.1 eg. woodland, managed grassland, unmanaged land and an area adjacent to fresh water. Within each site, the sample locations were chosen that represent both potential alibi sites and potential crime scenes such as recreational areas where a person could legitimately come into contact with earth materials, or secluded spaces and thoroughfares, which could provide opportunities for crimes to be committed.

Table 5.1: Visible and Land Use Characteristics of Sample Locations

Location	Description
1: Managed Grassland	A flat area of well-maintained, cut grass used for exercise and sporting activities.
2: Adjacent to Fresh Water	A flat area of miscellaneous wild vegetation, immediately adjacent to a fresh-water pond, housing various water fowl, and with restricted pedestrian access
3: Unmanaged Land	A natural footpath through a sloping area of miscellaneous wild vegetation, such as wild flowers and grasses.
4: Woodland	A natural footpath through a flat area of bare earth with a dense canopy of primarily deciduous trees, shrubs and localised leaf litter, immediately adjacent to a residential area and used as a thoroughfare to the park entrance.

The descriptors detailed in table 5.1 were applicable to each location, at each site, and photographs of each location type are shown in figures 5.1 to 5.4.

Figure 5.1 shows an overview of the vegetation and land use at each of the four sampling locations at the Brockwell Park site in London.

Figure 5.1: Sampling locations within Brockwell Park , London

Figure 5.1 Clockwise from top left: Location 1 (managed grassland), Location 2 (adjacent to fresh water), Location 3 (unmanaged vegetation), Location 4 (woodland).

Figure 5.2 presents an overview of the vegetation and land use at each of the four sampling locations at the Central Park site in New York City.

Figure 5.2: Sampling locations within Central Park, New York City.

Figure 5.2 Clockwise from top left: Location 1 (managed grassland), Location 2 (adjacent to fresh water), Location 3 (unmanaged vegetation), Location 4 (woodland).

Figure 5.3 illustrates a generalised view of the vegetation and land use at each of the four sampling locations at the Craigiebuckler Estate in Aberdeen

Figure 5.3: Sampling locations within Craigiebuckler Estate, Aberdeen

Figure 5.3 Clockwise from top left: Location 1 (managed grassland) Location 2 (adjacent to fresh water), Location 3 (unmanaged vegetation), Location 4 (woodland)

Figure 5.4 presents the general perspective of the planting and land use at each of the four sampling locations at the Lochend Park in Edinburgh

Figure 5.4: Sampling locations within Lochend Park, Edinburgh

Figure 5.4 Clockwise from top left: Location 1 (managed grassland), Location 2 (adjacent to fresh water), Location 3 (unmanaged vegetation), Location 4 (woodland).

5.2.1 Sample Collection and Preparation

At each of these locations, samples were taken from areas of exposed soil, which were deemed to be easily transferable and therefore forensically relevant. Five samples were collected from each location in order to assess intra-location variability, using the grid suggested for sampling footprints and tyre tracks by Pye (91), as outlined by McCulloch et al. (33) (34) (35). In accordance with Simmons (168), samples were gathered using a stainless steel spatula, removing any turf or gravel, where present. Approximately five grams of topsoil (<1cm depth) was collected at the corners and central point of a 1m square grid.

250mg of dry soil was added to a 1.5ml sterile, DNA free, polypropylene centrifuge tube and 0.5ml gradient grade acetonitrile was added by pipette. The tubes were placed in a sonic bath for 20min then centrifuged for 15 min at 13,000rpm. The supernatant was then passed through a 0.22µm PTFE syringe filter into an HPLC vial.

The samples discussed in this chapter were sampled at different times and locations to ensure that the marker peaks selected were representative of soil profiles at different times of year and across regions of varying geographies, discussed further in Chapter 6.

The initial data analysis approaches presented here used the results generated from the feasibility study conducted in London and an early iteration of the HPLC parameters, (Chapter 3 section 3.4.1), while the later strategies used in this chapter were applied to samples collected across all four sites and the final version of the HPLC method (Chapter 3 section 3.4.2).

5.2.1.1 Initial Instrument Parameters

The initial data analysis approaches presented here used the results generated from the feasibility study (Chapter 4) conducted in London and an early iteration of the HPLC parameters. 50 µl of each sample was injected onto a Shimadzu VP series HPLC system, which was comprised of an SIL-10AD autosampler with cooling tray, a FCV-10AL solvent mixing system, a DGU-14A vacuum degasser, a SCL-10A control unit, a SPD-M10A diode array detector (DAD) and a CTO-10AS column oven. The column used was a Whatman Partasil 10 ODS analytical column (10um particle size, C18 packing material, column dimensions 250 x 4.6mm) held at 25°C in order that fluctuations in ambient temperature did not affect reproducibility. The DAD collected UV-Vis spectra from 190-800nm with monitoring at 254nm

and sample vials were held at 10°C during analysis to minimise sample degradation during analysis. Gradient elution was used in order to speed up elution of strongly retained compounds and improve the practicality of the method, the gradient is detailed in Table 5.2, below. UHQ water and HPLC Gradient grade Acetonitrile were used and equipment was subject to a Good Clinical Laboratory Practice maintenance schedule.

Table 5.2: HPLC Gradient for Initial Analysis

Time (minutes)	% Water	% Acetonitrile
0	45	55
5	45	55
35	75	25
38	98	2
40	98	2
45	45	55
50	45	55

Samples were injected in a random order based on their identification number in order to prevent any systematic instrumental bias. Integration of the chromatograms was performed by the Shimadzu LC Solution Data Analysis software package, and peaks were detected and identified on the basis of predetermined acceptance criteria to ensure a consistent approach was used for all samples.

5.2.1.2 Final Instrument Parameters

A series of method development experiments (See Chapter 3) were first performed on the HPLC method used in previous studies to yield useful, discriminatory profiles. From these experiments, the optimum column, mobile phase and gradient parameters required to maximise the number of peaks detected per run and to reduce the overall sample analysis time were determined. Table 5.3 details the instrument parameters selected for use following the method development experiments. Samples were injected onto an Agilent 1100 HPLC system with DAD detector, using UHQ water as mobile phase A and gradient grade acetonitrile as mobile phase B, which had been degassed by sonication prior to use.

Table 5.3: Final HPLC Parameters

Injection volume	50µl		
Column	Waters Xbridge C18, 3.5um, 150x4.6mm at 30 ^{o}C		
Gradient	Time (min)	% Mobile Phase A	% Mobile Phase B
	0.0	53	47
	3.0	45	55
	24.0	26	74
	29.0	2	98
	31.0	2	98
	32.0	53	47
	35.0	53	47
Flow Rate	1ml/min		
Detector Settings	254nm, bandwidth 4nm, peak width >0.1min		

5.2.2 Data Analysis

The methods used to analyse the data are provided in section 5.2.2.1 for visual comparison and in section 5.2.2.2 for the statistical analysis.

5.2.2.1 Visual Comparison

The previously reported studies (18) (31) (30) into the use of HPLC as a tool for forensic geoscience have utilised visual inspection as the primary approach to data analysis, therefore visual analysis strategy was implemented as the initial strategy for this research.

Visual Comparison of Chromatograms

HPLC data is typically assessed by visual examination of the chromatography, and it is often fairly easy to identify groups of samples by overlaying the data from all samples on the same chromatogram. The initial data analysis technique was, therefore, a simple visual comparison of the sample chromatograms, to look for distinctive features in the chromatography and attempt to classify the anonymised samples based on these features.

The initial data set was then subject to further, more detailed visual examination to attempt to identify which peaks, if any, were potentially useful markers for comparing the groups of samples.

Visual Comparison of Retention Times only

Since pre- syn- and post- forensic event mixing and sample degradation can affect relative peak heights, it would be ideal if soil obtained from known locations could be grouped

together solely on the basis of shared peaks, identified by their retention time, and that the sets of peaks present in each group were distinct from one another. All peaks above the limit of detection (LOD), which was set at a minimum peak signal: noise ratio of 3:1, were therefore plotted (Figure 5.6) for each sample in the initial data set in an effort to compare samples based solely on the presence and absence of peaks.

Visual Comparison of Peak Heights and Retention Times

Early research into HPLC profiling of soils achieved some success in discriminating different sites using quantitative differences between the chromatograms, based not only on the retention times of the peaks, but also the relative sizes of the peaks, (18) (30) (32) (31) therefore this approach to data analysis was also used on the data generated in the feasibility study. It was not possible to overlay the chromatograms generated using the chromatography data system available at the time of the initial analyses, therefore the retention time and height of the apexes of all peaks above the LOD were plotted for the anonymised, initial sample set (Figure 5.7) using Microsoft Excel. The height of each peak was first adjusted to account for differences in sample concentration then the plot was examined for the presence of location-indicating peak clusters or peak profiles that were consistent within a particular group, and consistently different between groups.

5.2.2.2 Statistical Analysis

The visual classification process was naturally subjective and very time consuming due to the nature of the complexity of the data. In order to help reduce the subjectivity and the potential risk of cognitive bias during data analysis, and to quantify the significance of the discrimination between sample groups, further statistical analyses were performed on the data obtained from samples which had been analysed using the final HPLC instrument parameters selected in the method development section (Chapter 3).

Full Data Set

The HPLC profiles were examined using Agilent Chemstation software, and were also found to contain hundreds of individual, closely eluting peaks, many of which were close to the limit of quantification (LOQ), defined as peaks with a signal to noise ratio of 10:1. All data was first adjusted for variations in sample concentration, then the peaks below LOQ were removed from the dataset as they cannot be accurately quantified by the instrument.

After integrating and processing the chromatographic data, Canonical Discriminant Function Analysis(CDFA) was performed on the data using Statistical Package for Social Scientists (SPSS) to determine the accuracy and precision with which different sets of HPLC peaks allow samples to be grouped according to their location within each site. This data analysis technique used

the HPLC peaks as predictor variables, and each sample location as a grouping variable. The data for each individual sample were used by the software to generate functions, which are linear combinations of the variables that maximise the difference in the canonical discriminant function scores for each location. Each sample was then plotted using the scores for each function as co-ordinates, to create a scatter plot where samples of similar composition clustered closely together allowing groups of samples, and the relative degree of difference between groups, to be visualised. The functions were then used to assign each sample in the dataset to a particular location, based on the scores for each function, and the accuracy of classification was determined by comparing the predicted sample location to the true sample location.

CDFA was first performed using each of the chromatographic peaks observed in the dataset as variables in this analysis, on the data from samples collected at the London site (Figure 5.8), which were collected as part of the feasibility study (Chapter 4) then repeated for the samples collected in London in June 2013 (Figure 5.9) and November 2013 (Figure 5.10), which were collected as part of the study into seasonal variability (Chapter 6).

Peak Sub-Sets

Full datasets were prepared in the same way for the chromatography obtained at the Edinburgh and Aberdeen sites in January 2014, however it was decided at this stage, that it was necessary to reduce the number of peaks used in the data analysis, since there were up to 114 individual chromatographic peaks in each data set. Great care was required to ensure that like with like comparisons were being made between chromatograms, since there were so many peaks, many of which were very close in retention time, and as a result the length of time required to prepare the full datasets for analysis by CDFA was prohibitive to routine use of this technique.

Two smaller subsets of peaks were therefore chosen to use as markers in further analyses (table 5.4), in order to simplify the data analysis, improve efficiency and to reduce error rates: The first subset of peaks (set A) contained the 20 largest peaks observed in the three full datasets prepared for the London, Aberdeen and Edinburgh sites, since these were the easiest peaks to unambiguously identify and quantify and, as the major components of the sample, were the best targets for further method optimisation and sample size reduction; The second set (set B) were selected using the R “Subselect” package to identify smaller sets of variables which give equally good classification accuracies as the full data set, based on Wilks’ Lambda. This process produced ten sets of three or four peaks which, consisting of eleven individual

peaks in total, when used as variables in a leave-one-out classification, have near-perfect error rates when classifying the samples.

Table 5.4: Retention Times of Marker Peaks

Marker Set	Peak Retention Times
A	4.4, 9.0, 9.4, 10.0, 10.8, 11.6, 12.2, 12.6, 13.6, 14.2, 15.0, 15.5, 15.8, 18.8, 19.6, 20.3, 23.6, 24.3, 37.3, 30.4, 30.8
B	1.9, 4.4, 6.7, 12.2, 13.2, 13.7, 15.0, 19.1, 24.5, 26.9, 28.5

After identifying the peaks of interest in R, the data for both sub-sets of peaks were excised from the larger datasets from each of the four locations at each of the four sites sampled in January 2014. These markers were then plotted using Microsoft Excel for all samples, in order that the regions of variability in the profiles could be more easily visualised, and subsequently used as variables in the CDFA (Figures 5.11-5.18).

5.3 Results

The results obtained using all of the peaks observed in the chromatograms are presented in section 5.3.1 and the results using the peak marker sets are provided in section 5.3.2

5.3.1 Full Data Set

The profiles used for visual comparison of the samples using all of the peaks detected in the chromatograms for each data set are displayed in figures 5.5-5.7 and in table 5.5, and the results of the CDFA for these data are presented in figures 5.8-5.10 and table 5.6.

5.3.1.1 Visual Comparison

Typical chromatograms obtained for the anonymous samples in the initial analysis are displayed in figure 5.5, and the results of the visual comparison and classification are presented in table 5.5

Figure 5.5 Visual comparison of chromatograms

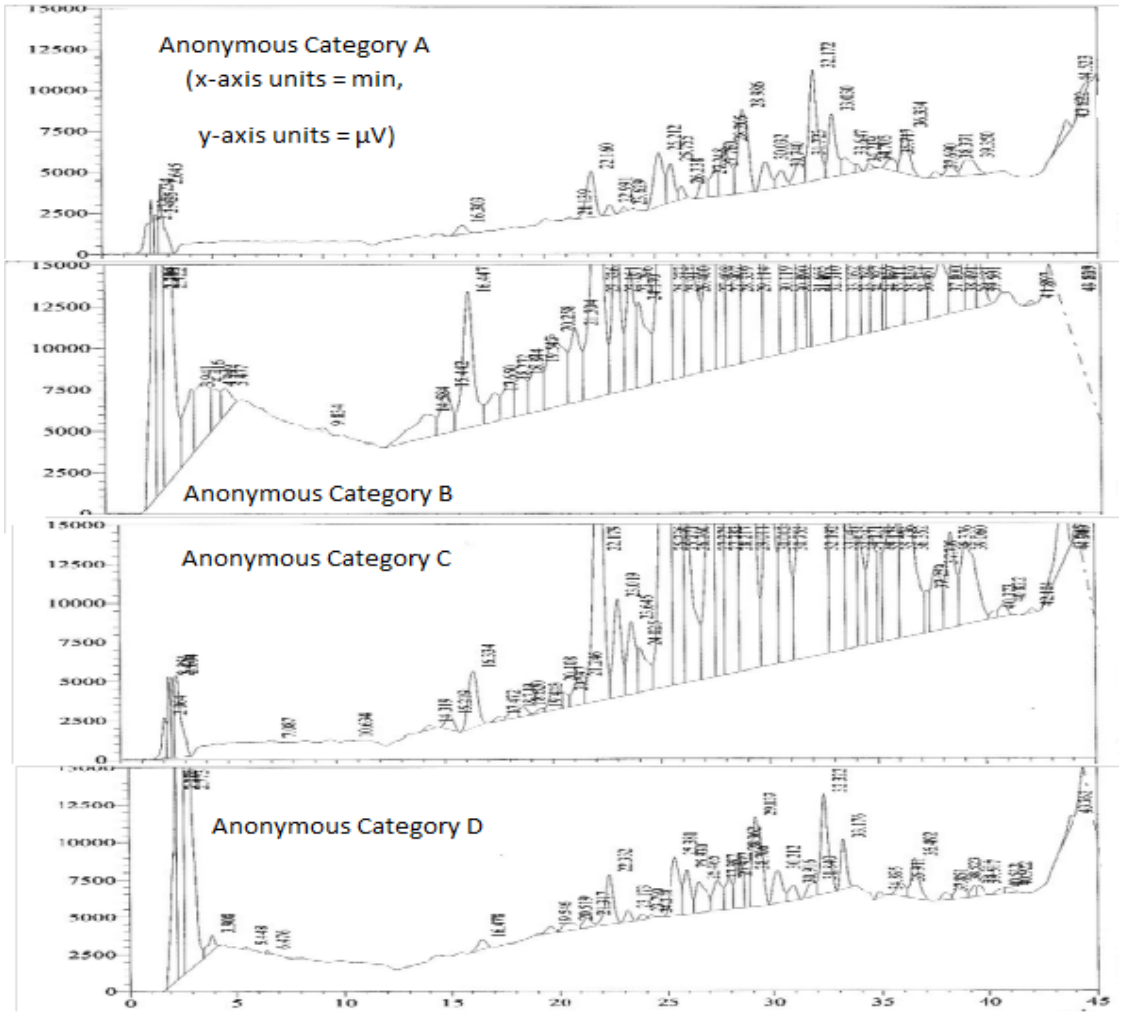


Figure 5.5 Visual comparison of chromatograms from feasibility study (Chapter 4)

Table 5.5 provides details of the criteria used to classify the anonymous samples, and lists the samples assigned to each category.

Table 5.5: Classification of samples by visual comparison of chromatograms

Anonymous Category	Anonymised Samples Present	Sample Points (respectively)	Major Peaks Present		
			3-5min	17-18min	20-21min
A	5, 8, 14, 18, 19	1E, 1B, 1D, 1C, 1A	No	No	No
B	1, 2, 4, 9, 10	4E, 4B, 4D, 4A, 4C	Yes	Yes	Yes
C	6, 7, 9, 13, 20	2D, 2A, 2C, 2B, 2E	No	Yes	Yes
D	3, 11, 15, 16, 17	3C, 3E, 3A, 3B, 3D	Yes	No	Yes

Four distinctive chromatographic profiles were observed upon systematic examination of the chromatograms of anonymised samples (Figure 5.5). Combinations of characteristic baseline features and groups of peaks were common to samples taken from the same location, which allowed samples to be grouped with 100% accuracy (table 5.5).

Figure 5.6: Plot of Sample Groups by Retention Time

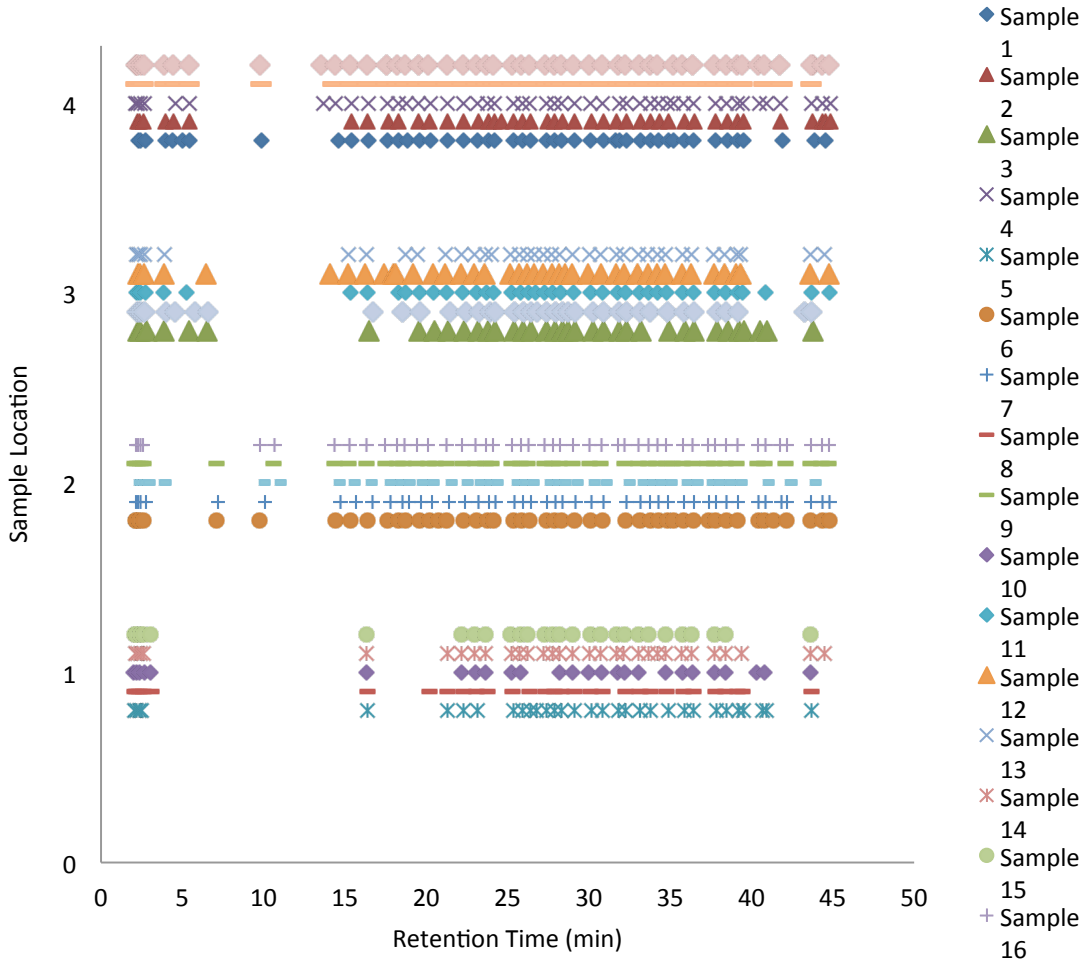
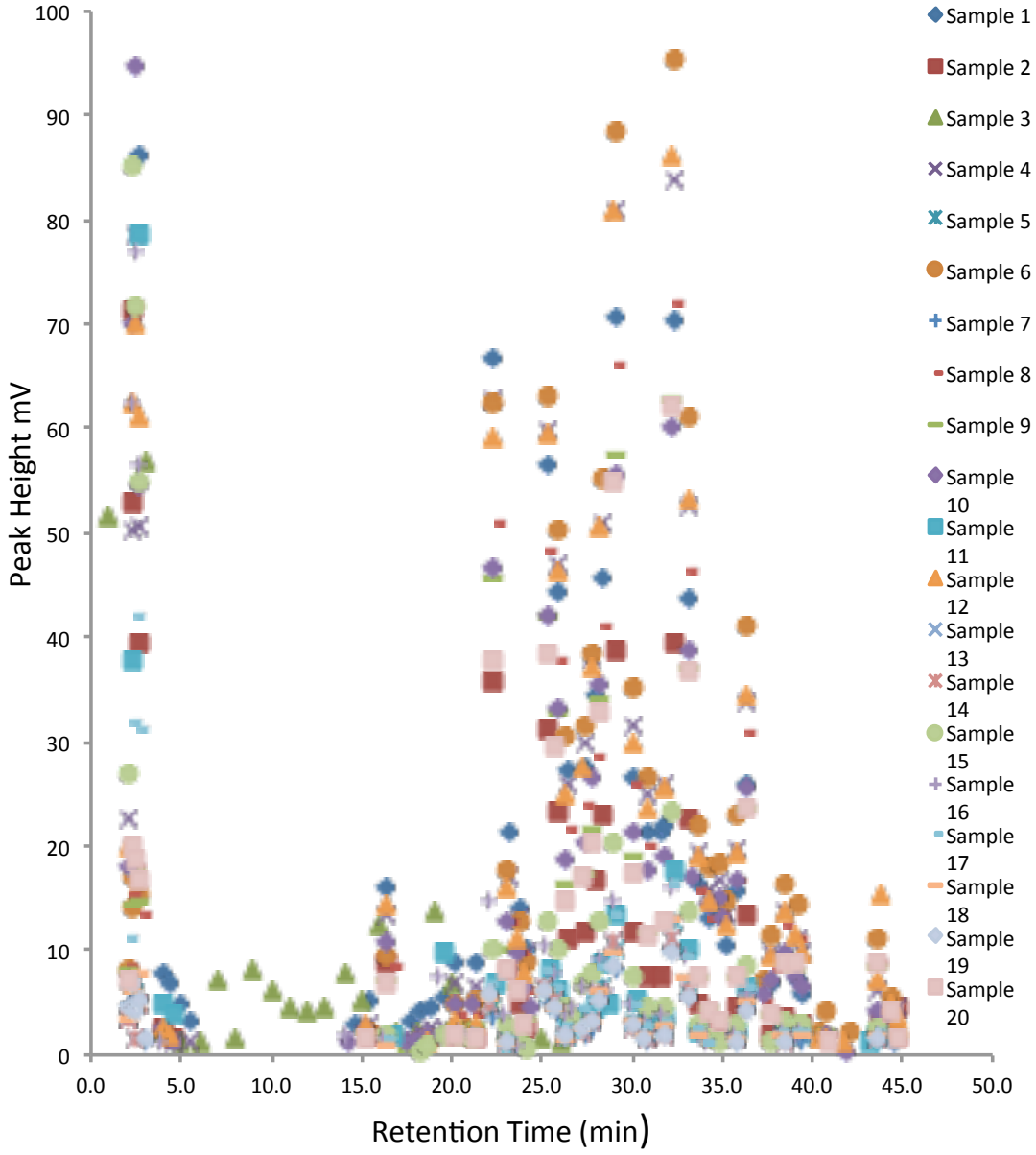


Figure 5.6: Plot of Sample Groups by Retention Time

The plot of the observed retention times displayed in figure 5.6, shows that it is not possible to exclusively group samples on the basis of the retention times of their constituents. Samples from the same location showed a similar pattern of retention times, however each sample at a location did not share all the same peaks as other samples from that location. None of the locations displayed common set of peaks that was entirely distinct from the other groups. Figure 5.6 shows that each Location displayed internal variability in terms of the number and retention time of peaks present, therefore it was not possible to categorise samples in absolute terms on this basis. It is possible that inherent variation in the HPLC's pumping efficiency could have led to changes in retention times between samples, leading to inaccurate identification of peaks. This could be avoided in future analyses through the addition of an internal standard, allowing components to be identified on the basis of their retention times relative to the standard peak.

Figure 5.7: Plot of Sample Groups by Retention Time and Peak Height



No clear patterns were observable in figure 5.7, which displays the data on the basis of the retention time and height of the peak apexes, due to the number of samples and peaks, therefore samples could not be grouped on this basis. The number of peaks and the complexity of the chromatographic profiles obtained in this study meant that it was not only not possible to identify distinct groups of sample locations in this way, but it is also highly impractical as an approach to interpreting this type of data.

5.3.1.2 Statistical Analyses

The results of the statistical analysis for each of the sample sets by CDFA are presented in table 5.6.

Table 5.6 Summary of CDFA results using all peaks

HPLC Profiles	Classification Accuracy %	Wilks' Lambda Significance test of functions			% Variance Function 1	% Variance Function 2	% Variance Function 3
London		1-3 p=	2-3 p=	3 p=			
Summer 2013 (Initial method)	100.0	0.000	0.000	0.000	54.6	35.1	10.3
Summer 2013 (Final method)	100.0	0.000	0.001	0.090	84.9	13.5	1.5
Autumn 2013 (Final method)	95.0	0.003	0.360	0.526	93.4	4.6	2.0

Figures 5.8, 5.9 and 5.10 show the scatter plots generated in the CDFA, which represent the relative differences in the samples and explains the extent of discrimination between the sample groups in the analyses of the data from London in Summer 2013 using the initial HPLC method, London Summer 2013 using the final HPLC method, and London in Autumn 2013, respectively.

Figure 5.8 CDFA Scatter plot using all peaks, London Summer 2013 (Initial Method)

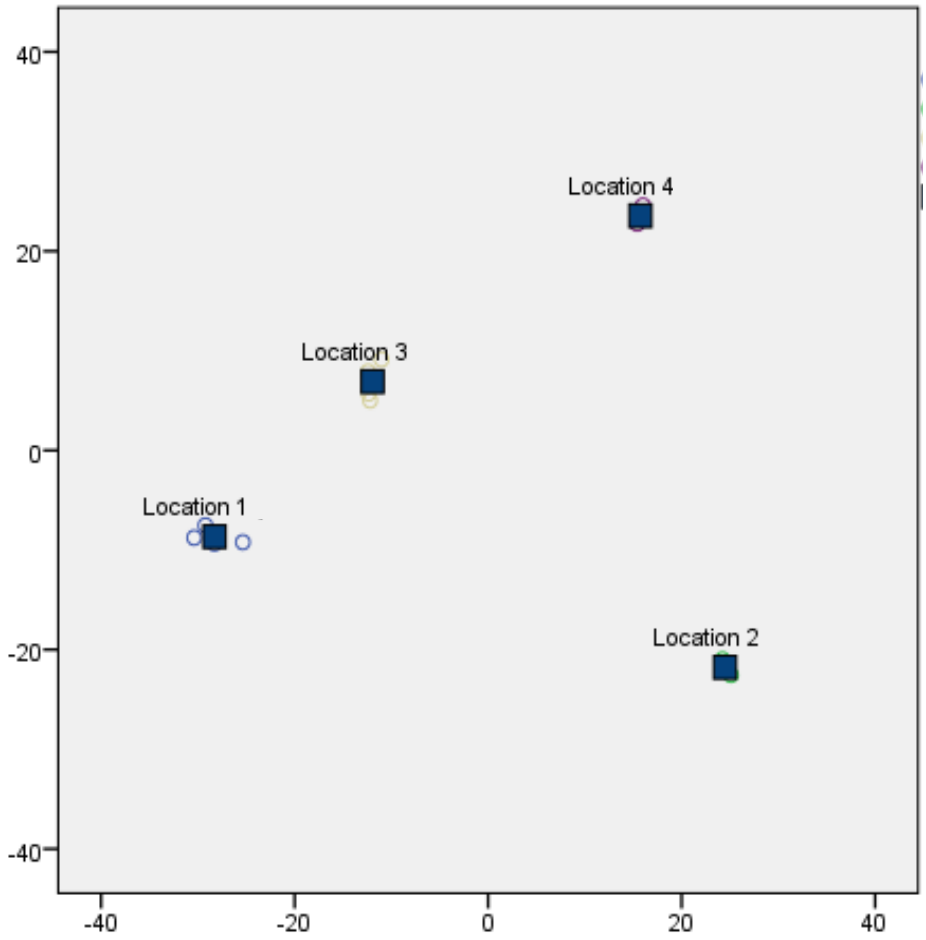


Figure 5.8 CDFA Results showing the scores for function 1 (x-axis) and function 2 (y-axis) using all peaks observed with the initial instrument parameters for samples collected from Brockwell Park, London, in Summer 2013

The results of the CDFA for the height of each of the peaks observed for the proof of concept samples analysed using the final HPLC method are shown in figure 5.8 and Table 5.6, above. 100% of samples were grouped correctly with these functions. The software determined three functions with which to categorise the samples, and functions 1, 2 and 3 explained 54.6%, 35.1% and 10.3% of the variance between the groups, respectively, and the discrimination between the groups of samples was significant at the 99% confidence level (p=0.000).

Figure 5.9 CDFA Scatter plot using all peaks, London Summer 2013 (Final Method)

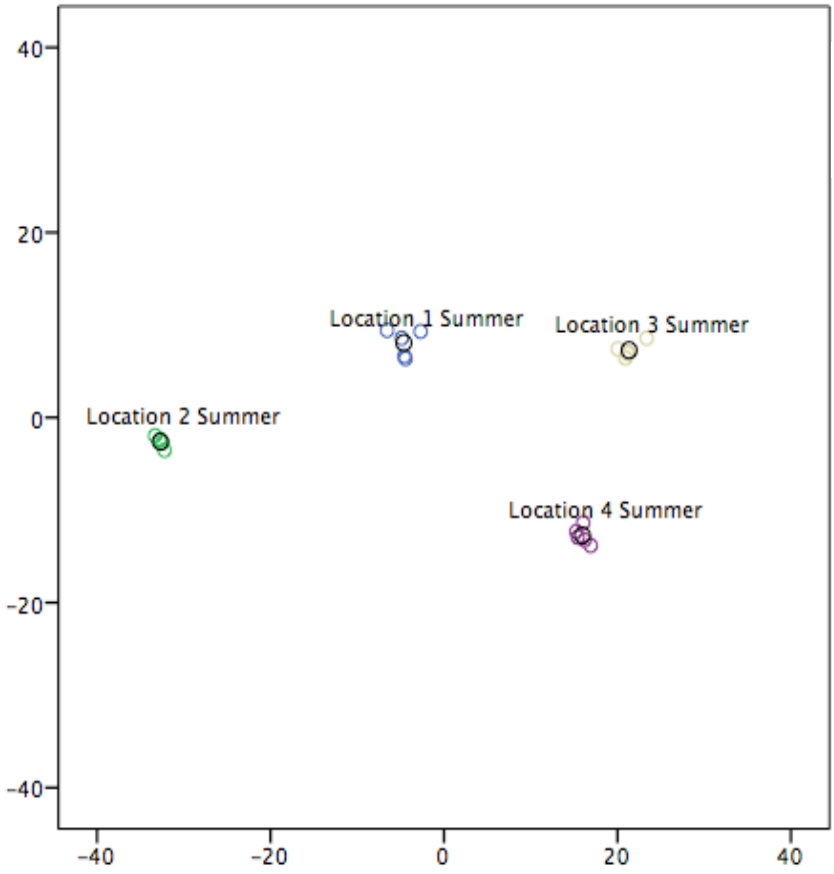


Figure 5.9 CDFA Results showing the scores for function 1 (x-axis) and function 2 (y-axis) using all peaks observed with the final instrument parameters for samples collected from Brockwell Park, London, in Summer 2013

The results of the CDFA for the height of each of the peaks observed for the samples collected in June 2013 are shown in figure 5.9 and table 5.6, above. 100% of samples were grouped correctly with these functions. The software determined three functions with which to categorise the samples, and functions 1, 2 and 3 explained 84.9%, 13.5% and 1.5% of the variance between the groups, respectively, and the discrimination between the groups of samples was significant at the 99% confidence level (p=0.000)

Figure 5.10 CDFA Scatter plot using all peaks, London Autumn 2013 (Final Method)

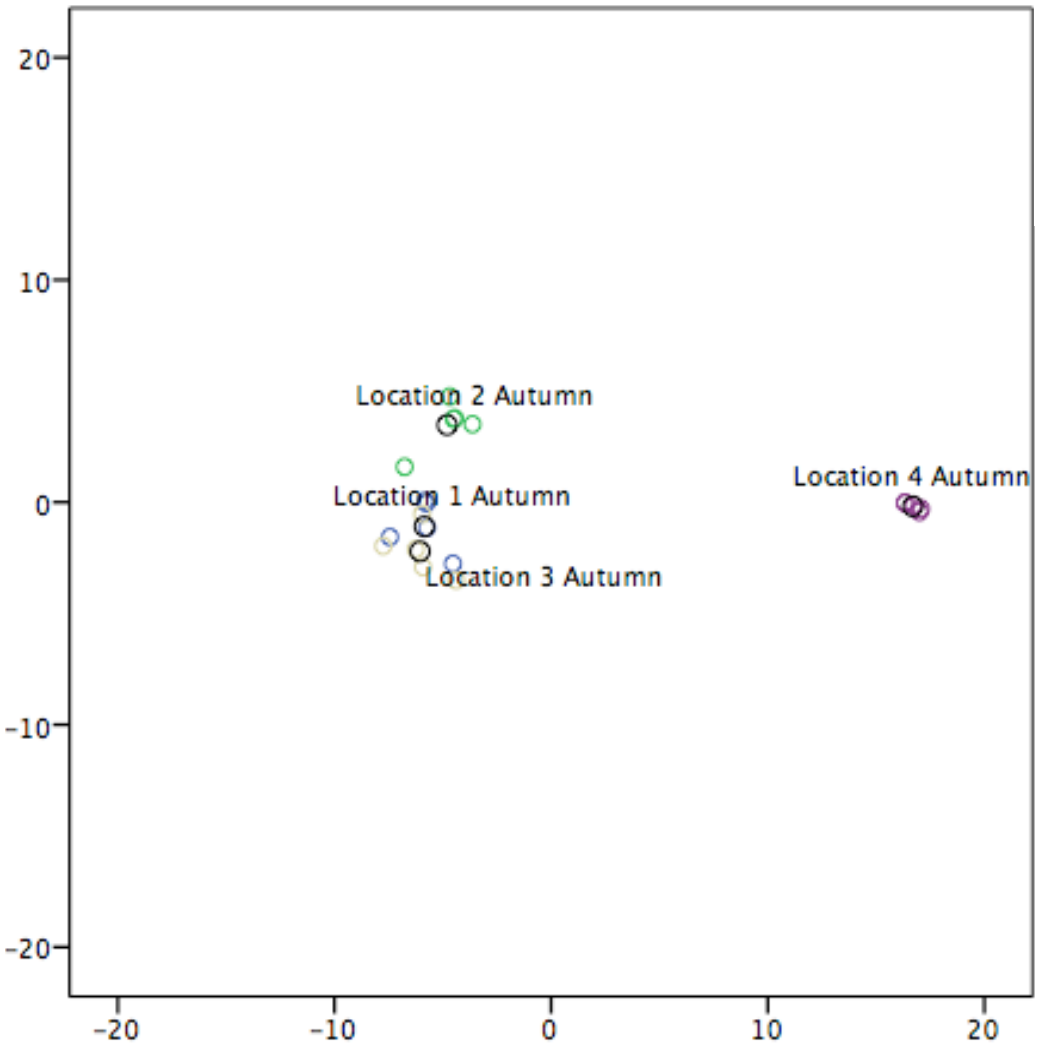


Figure 5.10 CDFA Results showing the scores for function 1 (x-axis) and function 2 (y-axis) using all peaks observed with the final instrument parameters for samples collected from Brockwell Park, London, in Autumn 2013

The results of the CDFA for the height of each of the peaks observed for the samples collected in November 2013 are shown in figure 5.10 and table 5.6, above. 95% of samples were grouped correctly with these functions, with one of the sample from unmanaged land was misclassified as having originated from managed grassland. The software determined three functions with which to categorise the samples, and functions 1, 2 and 3 explained 93.4%, 4.6% and 2.0% of the variance between the groups, respectively, and the discrimination of the groups was significant at the 95% confidence level (p=0.003)

5.3.2 Peak Subsets

The results obtained using the two marker peak sets are presented here, the discussion of the visual assessment is provided in section 5.3.2.1 and the results of the statistical analyses are displayed in section 5.3.2.2

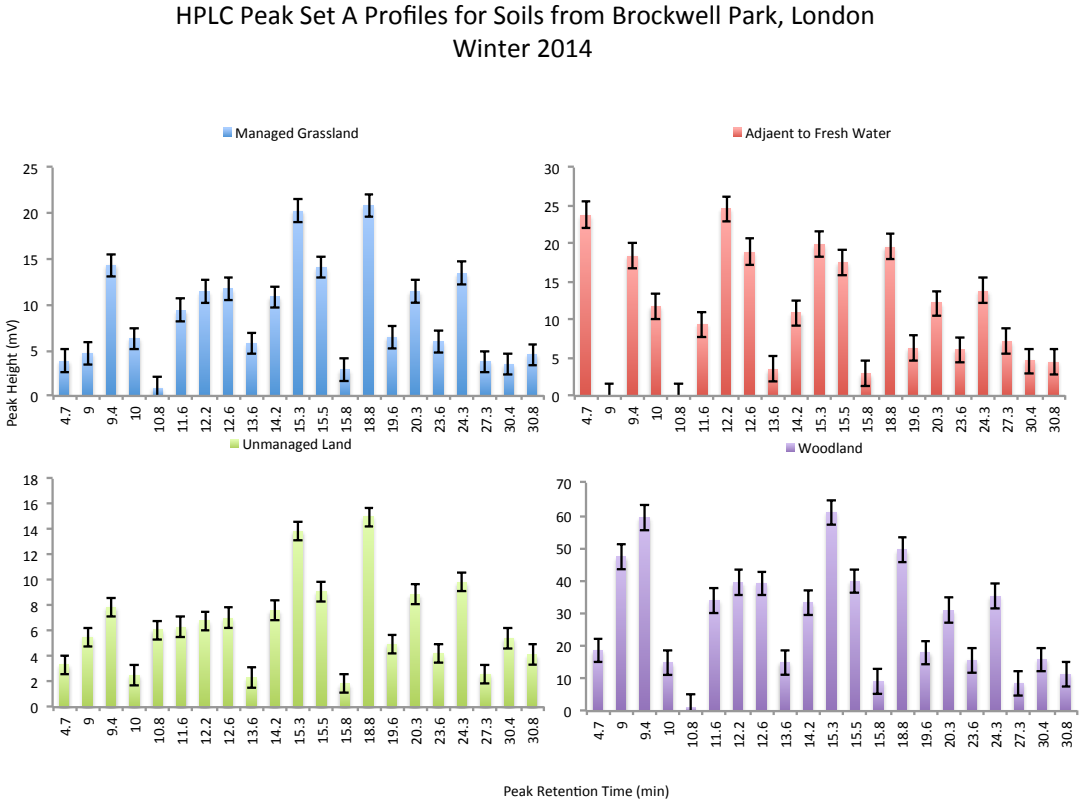
5.3.2.1 Visual Comparison

The profiles used for visual comparison of the samples using the two marker peak sets are illustrated in figures 5.11-5.18

HPLC Peak Set A

Figures 5.11-5.14 show the profiles obtained for HPLC peak set A for the London, Edinburgh, Aberdeen and New York City sites, respectively.

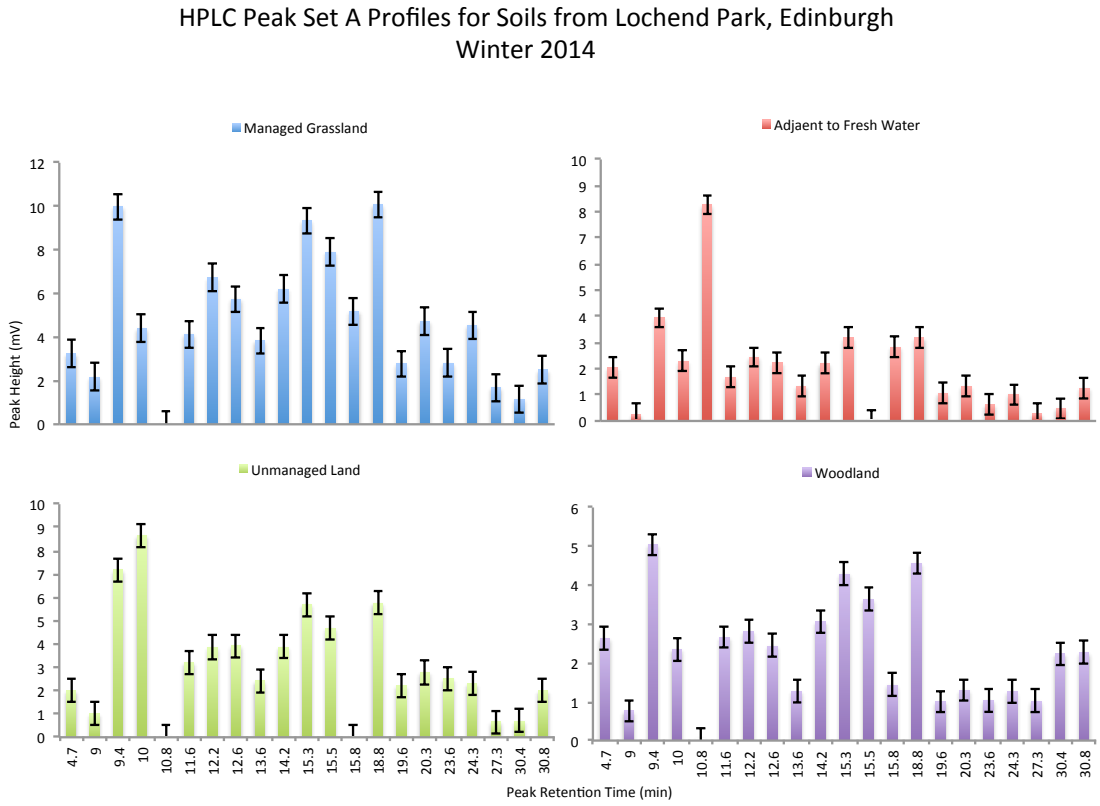
Figure 5.11: HPLC Profiles for Brockwell Park, London- Peak Set A



All four locations within Brockwell Park, London could be distinguished by the profiles of HPLC peak set A (Figure 5.11). The size order for the peaks at 4.7, 9.4 and 10min were distinct for samples adjacent to fresh water, as was the absence of the peaks at 9 and 10.8min. Managed grassland could be distinguished from woodland and unmanaged land by the size of the peak at 9.4min, the ratio of this peak compared with the peak at 9min was larger for managed

grassland, at 3:1, than for both unmanaged land and woodland, at 1.3:1 and 1.4:1, respectively. The profiles for woodland soils were noticeably different from the other locations as the magnitudes of the woodland peaks were more than double those of the other sites, while the relatively high ratio of 2.5:1 for the peak at 15.3min, relative to the peak at 9min, for woodland samples separated these samples from unmanaged land where the ratio was 1.2:1. Comparison of the two tallest peaks for each location was also useful in discriminating the samples. For soils adjacent to fresh water the two tallest peaks were 4.7 and 12.2min, whereas for woodland soils the peaks at 9.4 and 15.3min were largest, on the other hand the peaks at 15.3 and 18.8min were largest for both managed grassland and unmanaged land, suggesting the possibility that these peaks are indicative of grassland soils.

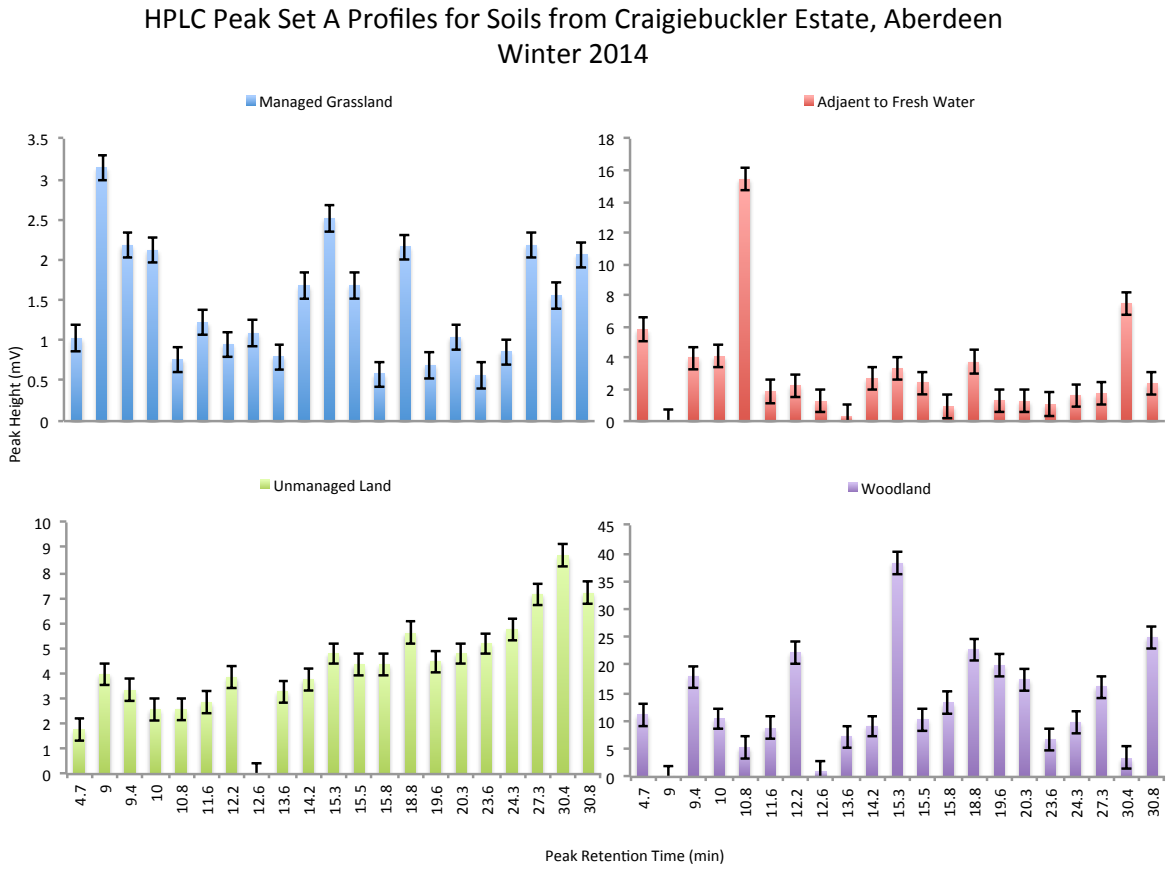
Figure 5.12: HPLC Profiles for Lochend Park, Edinburgh- Peak Set A



It was possible to discriminate all four locations within Lochend Park, Edinburgh on the basis of the profiles of HPLC peak set A (Figure 5.12). The presence of a peak at 10.8min was a useful discriminator for the soils adjacent to fresh water, since this was absent at the other three locations. The soils from unmanaged land were the only samples that did not contain a peak at 15.8min. The profiles obtained for woodland and managed grassland samples were arguably the most similar at this site, however several peak ratios were notably different at the two locations. At the managed grassland location, the ratio of the peak height at 9min relative

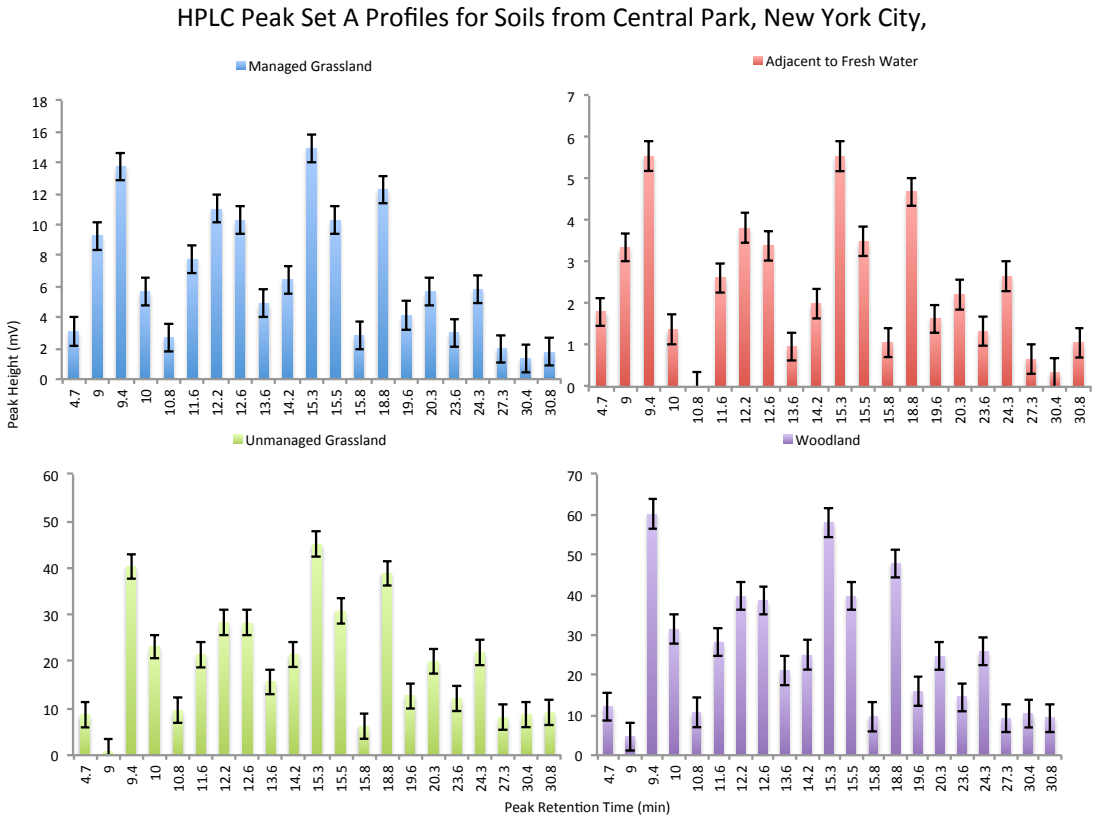
to the peak at 4.7min was much larger, at 3.5:1, than for woodland samples, where this ratio was 1.5:1, likewise the peaks at 20.3 and 24.3min were 69% and 63% larger for managed grassland than their neighbours at 19.6 and 23.6min, respectively, but only 30% and 26% larger for woodland soils. In addition, the peak at 30.4min was approximately equal in size to the peak at 30.8min for woodland soil, while for managed grassland the peak at 30.4min was half the size of the peak at 30.8min.

Figure 5.13: HPLC Profiles for Craigiebuckler Estate, Aberdeen- Peak Set A



The profiles for HPLC peak set A allowed each of the four locations in Craigiebuckler Estate, Aberdeen to be distinguished visually (Figure 5.13). The largest peak present in the managed grassland samples was the peak at 9.4min, which was distinct from the other three locations. The profiles of soils adjacent to fresh water differed from the other locations in that the largest peak was at 10.8min. The profiles were also distinct from the managed grassland and unmanaged land with the absence of a peak at 9min. The large relative height of the peak at 30.8min was distinctive of woodland soil profiles, and these profiles were also discernable from the unmanaged land and managed grassland through the absence of the peak at 9min. The profile of the samples from unmanaged land were distinctive with the highest peak at 30.4min, the presence of the peak at 9 min and the absence of a peak at 12.6min.

Figure 5.14: HPLC Profiles for Central Park, New York City- Peak Set A

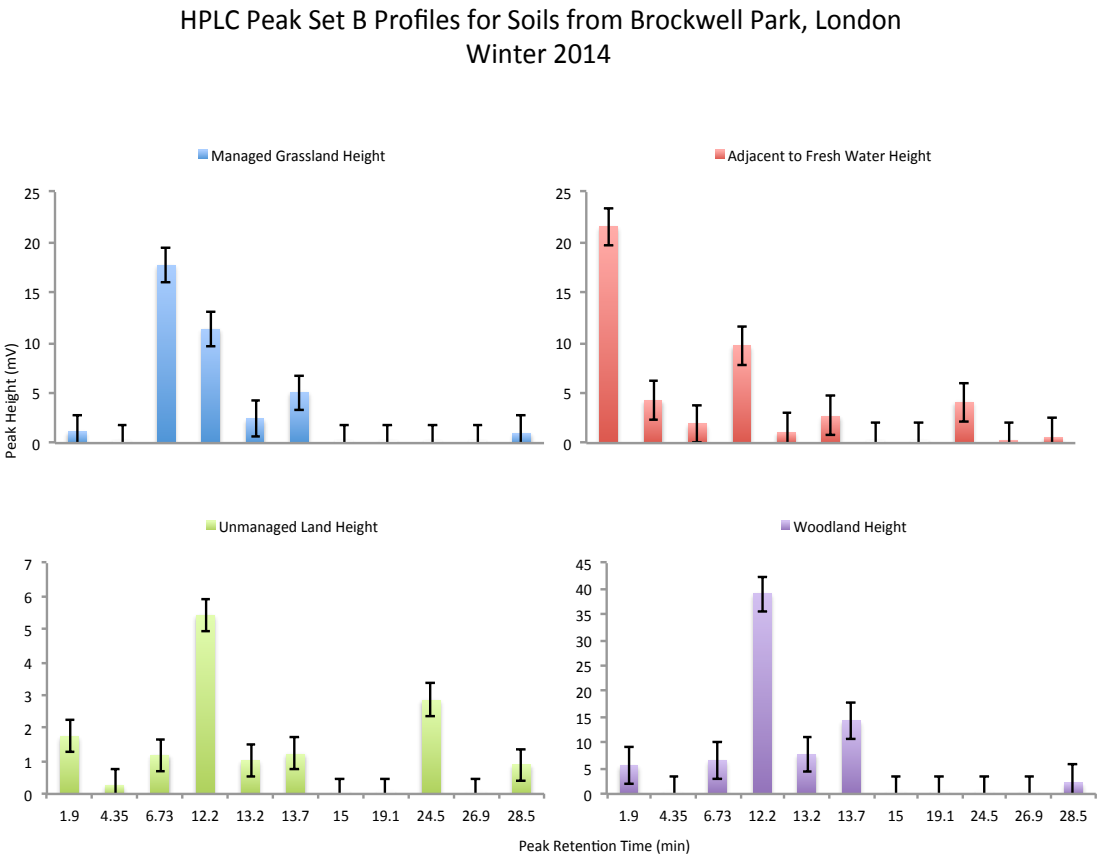


In Central Park, New York City, all four locations could be discriminated on the basis of their HPLC profiles for peak set A (Figure 5.14). The location adjacent to fresh water could be unambiguously distinguished from the other locations through the absence of a peak at 10.8min, while the absence of the peak at 9min was a unique feature of the soils from unmanaged land at this site. The profiles of the managed grassland and woodland locations were very similar, however there was a noticeable difference in the size of the peak at 9min compared to its neighbours at 4.7 and 9.4min between the two locations. The peak at 9min was larger than the peak at 4.7min, at 9mV and 3mV, respectively, for the managed grassland, while for the woodland location the peak at 4.7min was larger at 12mV compared to 5mV at 9min. In addition, the peaks were generally three times larger for the woodland soils, ranging from 5 to 60mV, than for managed grassland where the peaks ranged from 1 to 15mV.

HPLC Peak Set B

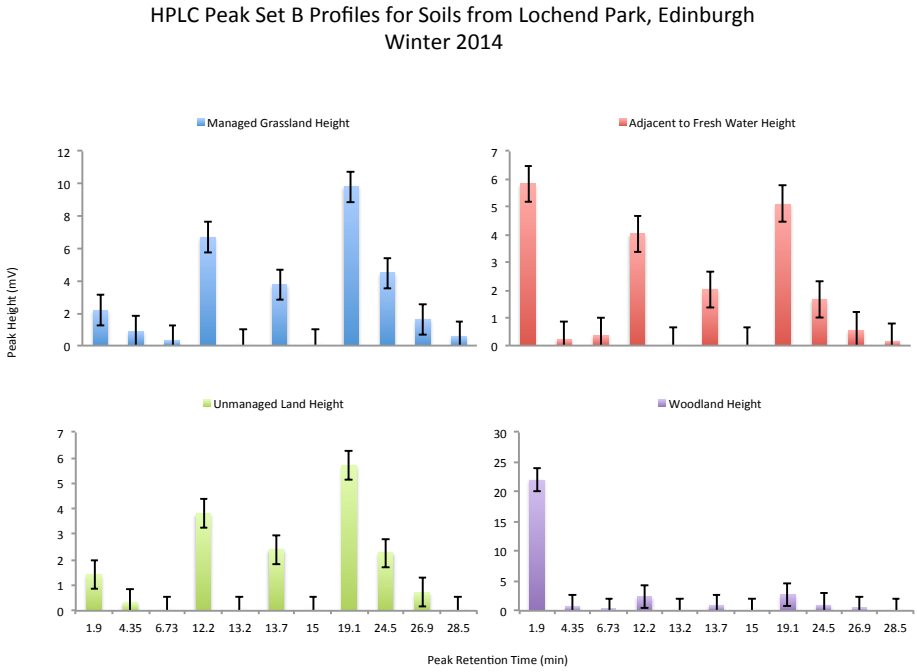
Figures 5.15-5.18 show the profiles obtained for HPLC peak set B for the London, Edinburgh, Aberdeen and New York City sites, respectively.

Figure 5.15: HPLC Profiles for Brockwell Park, London- Peak Set B



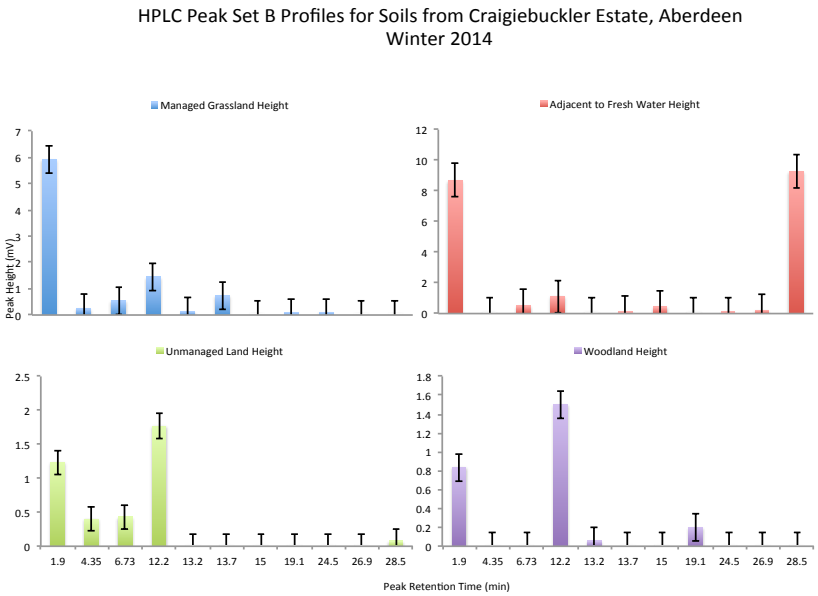
The HPLC profiles for peak set B allowed all four samples locations within Brockwell Park, London to be discriminated visually (Figure 5.15). The magnitudes of the peaks in woodland soils, with an average peak height of 39.0 mV for the largest peak at 12.2min, were approximately twice the size of those in soils adjacent to fresh water and managed grassland, where the largest peaks were 21.6 and 17.7mV, respectively and were in turn around three times the size of the peaks found in unmanaged land, where the largest peak was only 5.4mV. The large size of the peak at 6.73min relative to the peak at 12.2min distinguishes managed grassland from all other sample locations, while the large size of the peaks at 1.9min compared to all other peaks is distinctive of the profiles in soils adjacent to fresh water. The profiles of soils from unmanaged land and woodland were visually more similar, however on closer inspection the unmanaged land samples can be easily separated due to the presence of the peak at 24.5min which is absent in woodland samples, and interestingly, is not present in the soils from managed grassland therefore cannot be derived from grasses.

Figure 5.16: HPLC Profiles for Lochend Park, Edinburgh- Peak Set B



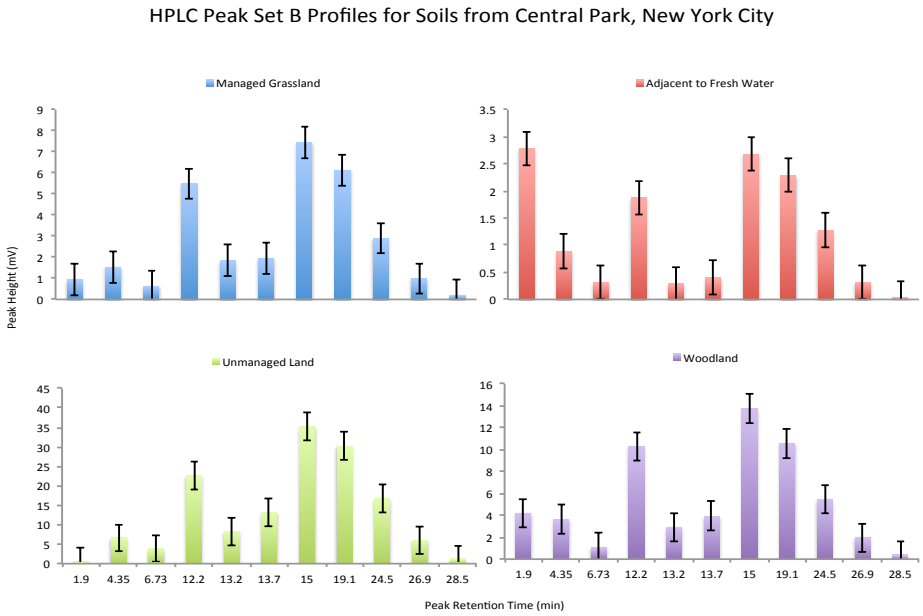
The profiles for HPLC peak set B were not as easily distinguishable at each of the four locations in Lochend Park, Edinburgh. Under scrutiny, however, each profile was different from the others (Figure 5.16). The woodland samples were distinctive in that they had large peaks at 1.9min, which, at 22.0mV, were approximately ten times the size of the next largest peaks at 12.2 and 19.1min at 2.4 and 2.8mV, respectively. The soils adjacent to fresh water were distinct with the two largest peaks at 1.9 and 19.1min, and these were also very similar in size to one another. It was more difficult to visually discriminate the profiles of managed grassland and unmanaged land, however the largest peak in the former was 9.8mV, approximately twice the size of the same peak in the latter, at 5.6mV. In addition, the small peaks present in the managed grassland profiles at 6.7 and 28.5min were absent in the samples from unmanaged land. However the relative uncertainty in measurements of peaks at this low level is far greater and therefore there can be much less confidence in the visual discrimination of these two locations using these low level peaks.

Figure 5.17: HPLC Profiles for Craigiebuckler Estate, Aberdeen- Peak Set B



The samples from the four locations within the Craigiebuckler Estate, Aberdeen were easily distinguished using peak set B (Figure 5.17). Comparison of the retention time of and ratio between the two largest peaks at each location was useful in grouping the samples from managed grassland, with the largest peak at 1.9min and next largest at 12.2min, and those from soils adjacent to fresh water, which had two large peaks of similar size, at 1.9 and 28.5min. The typical profiles of unmanaged land and woodland samples were similar in having the largest peak at 12.2min and the next largest at 1.9min, however the ratio of the former to latter was greater for woodland soils, at 8:1 compared to 3:1 for the unmanaged location, and woodland soils were also missing the peaks at 4.35 and 6.73min that were present at the unmanaged location.

Figure 5.18: HPLC Profiles for Central Park, New York City- Peak Set B



The profiles obtained for peak set B also varied across the four locations in Central Park, New York City (Figure 5.18). The peak sizes were far larger for unmanaged land and were smallest at the location adjacent to fresh water, while the peaks in the managed grassland samples were generally three times as large as those for the location adjacent to fresh water, and woodland samples were approximately twice the size obtained for managed grassland. Soil profiles for unmanaged land could be separated from the other three locations by the absence of the peak at 1.9min while samples from managed grassland could be distinguished from those for woodland by the size order of the peaks at 1.9, 4.35 and 6.73min, The profiles from soil adjacent to fresh water were most noticably different from the other locations in the ratio of the peak at 1.9min compared to the peak at 12.2 min, the peak at 12.2min was 1.87mV and was 33% smaller than the 2.78mV peak at 1.9min for soils adjacent to fresh water, while it was far larger than the peak at 1.9min for managed grassland with peak heighs of 5.49mV and 0.94mV, and for woodland the peak heights were 10.3mV and 4.19mV, respectively.

5.3.2.2 Statistical Analyses

Full details of the CDFA results are provided in Table 5.7 and the scatter plots produced for both peak sets are shown in Figure 5.19.

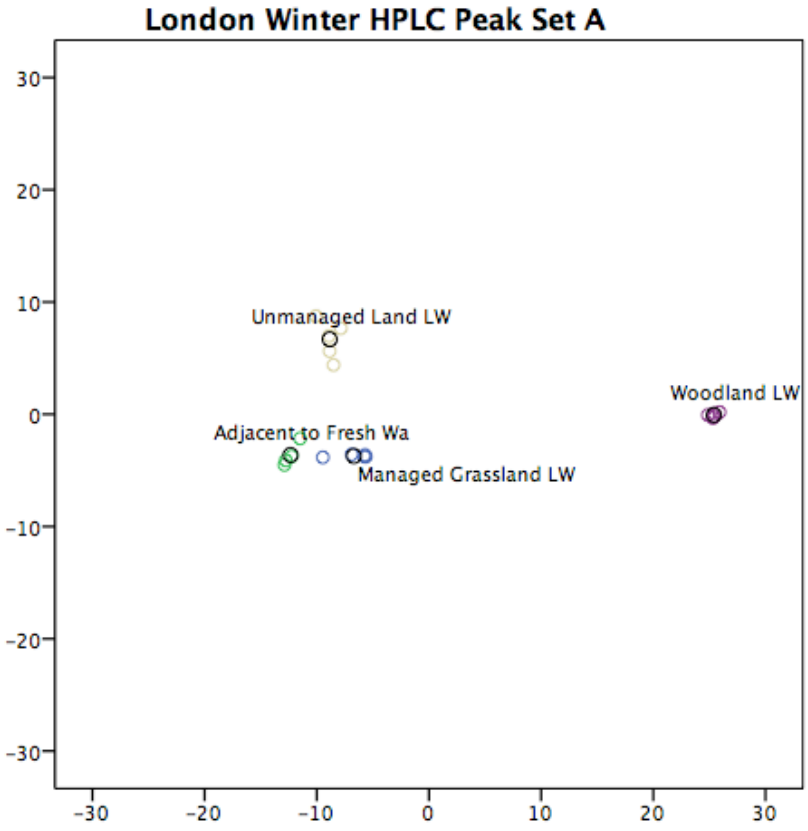
Table 5.7 Canonical Discriminant Function Results

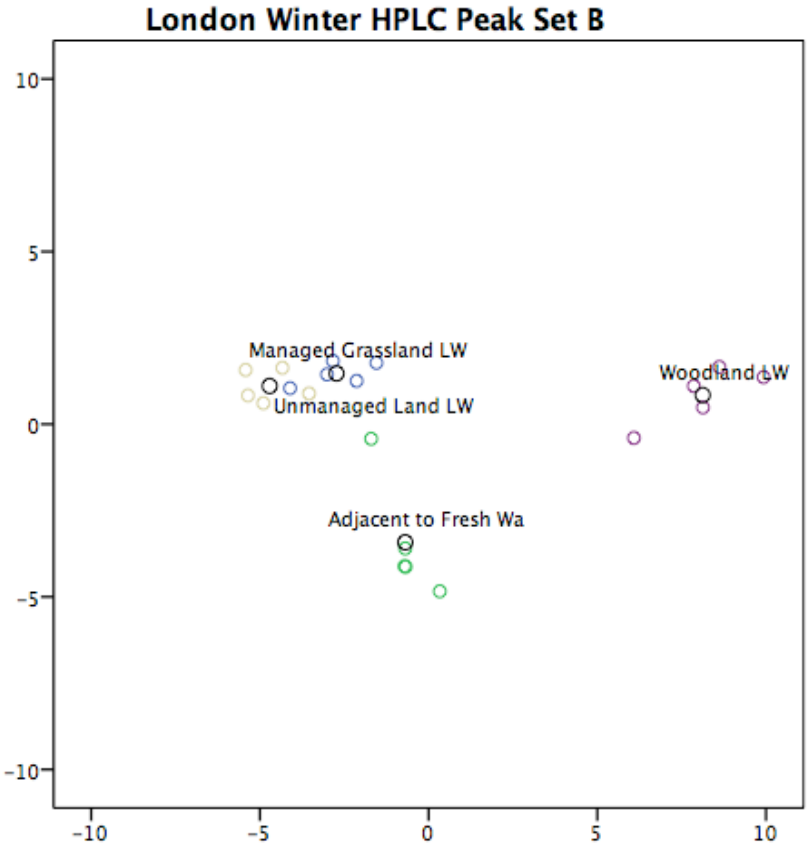
HPLC Profiles	Classification Accuracy %	Wilks' Lambda Significance test of functions			% Variance Function 1	% Variance Function 2	% Variance Function 3
London		1-3 p=	2-3 p=	3 p=			
A	100.0	.000	.002	.034	89.7	7.0	3.3
B	90.0	.000	.041	.684	84.8	13.9	1.3
Edinburgh							
A	100.0	.000	.000	.022	88.9	8.3	2.7
B	100.0	.000	.018	.397	62.4	32.5	5.0
Aberdeen							
A	94.7	.001	.147	.531	90.6	7.8	1.5
B	100.0	.000	.000	.014	97.4	2.4	0.2
New York							
A	100.0	.000	.000	.005	92.3	5.8	1.9
B	100.0	.000	.000	.071	73.3	24.1	2.5

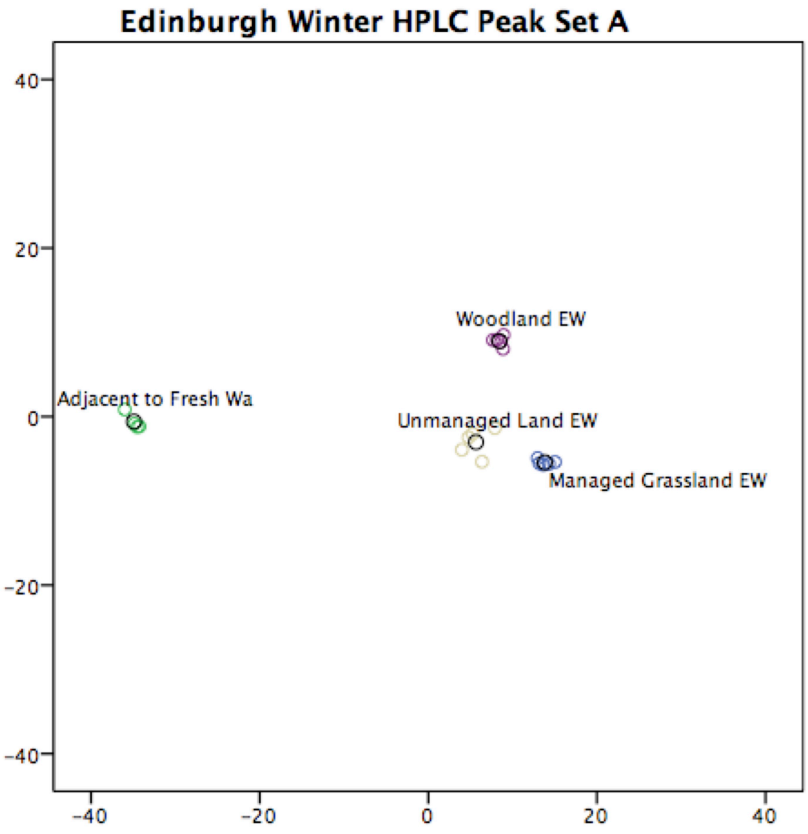
Using HPLC Peak Set A (Figure 5.19), all but one sample was classified to the correct location across all four sites. One sample from the unmanaged location in Craigiebuckler Estate, Aberdeen was misclassified as having originated from the woodland location, giving an overall accuracy rate of 94.7% for this peak set at this site (table 5.7). The discriminant functions gave rise to sample groupings that were significant at the 99% confidence interval ($p \leq 0.001$)

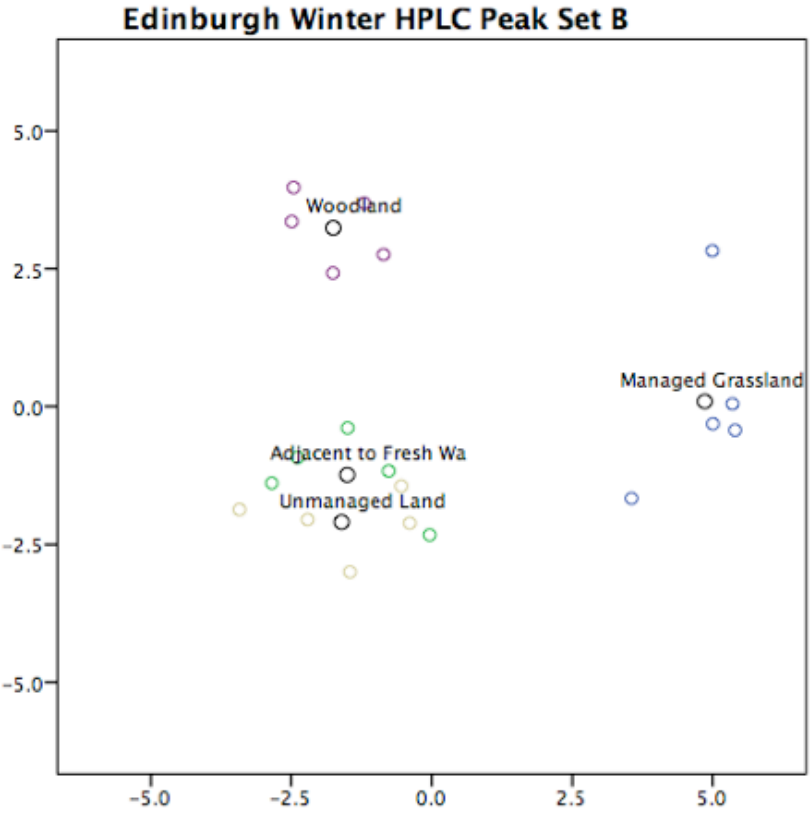
For HPLC Peak Set B (Figure 5.19), all the samples were correctly classified at the Aberdeen, Edinburgh and New York sites. Two samples from Brockwell Park, London were incorrectly assigned to groups using the functions generated. One sample from managed grassland was predicted to belong to the unmanaged land group, while one sample from the location adjacent to fresh water was incorrectly assigned to the managed grassland soil group. This resulted in a grouping accuracy rate of 90.0% for the London site (table 5.7). As with Peak Set A, the discrimination of sample groups resulting from the functions produced in this analysis, was statistically significant at the 99% confidence interval ($p = 0.000$).

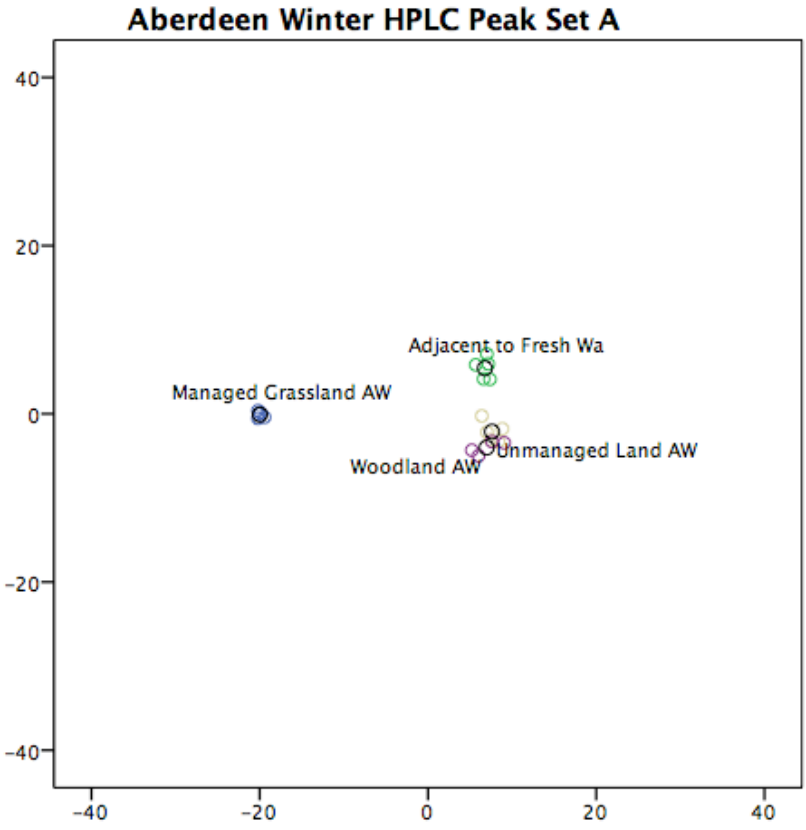
Figure 5.19: Canonical Discriminant Function Plots

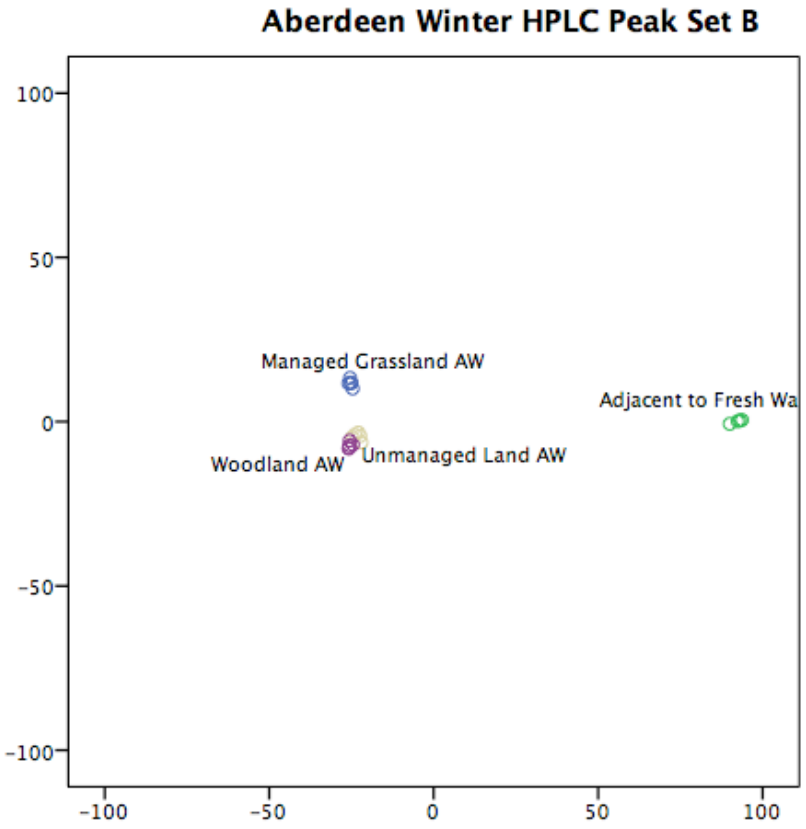


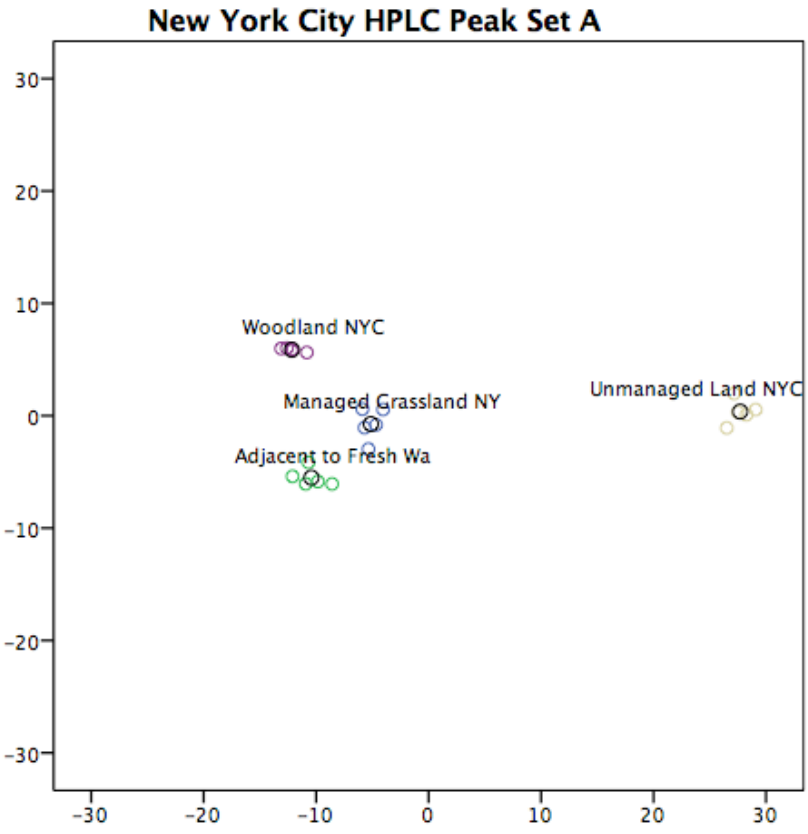



Edinburgh Winter HPLC Peak Set A
40
20
0
-20
-40
-40
-20
0
20
40
Woodland EW
Adjacent to Fresh Wa
Unmanaged Land EW
Managed Grassland EW


Edinburgh Winter HPLC Peak Set B
5.0
2.5
0.0
-2.5
-5.0
-5.0
-2.5
0.0
2.5
5.0
Woodland
Managed Grassland
Adjacent to Fresh Wa
Unmanaged Land


Aberdeen Winter HPLC Peak Set A
40
20
0
-20
-40
-40
-20
0
20
40
Adjacent to Fresh Wa
Managed Grassland AW
Unmanaged Land AW
Woodland AW


Aberdeen Winter HPLC Peak Set B
100
50
0
-50
-100
-100
-50
0
50
100
Managed Grassland AW
Adjacent to Fresh Wa
Woodland AW
Unmanaged Land AW



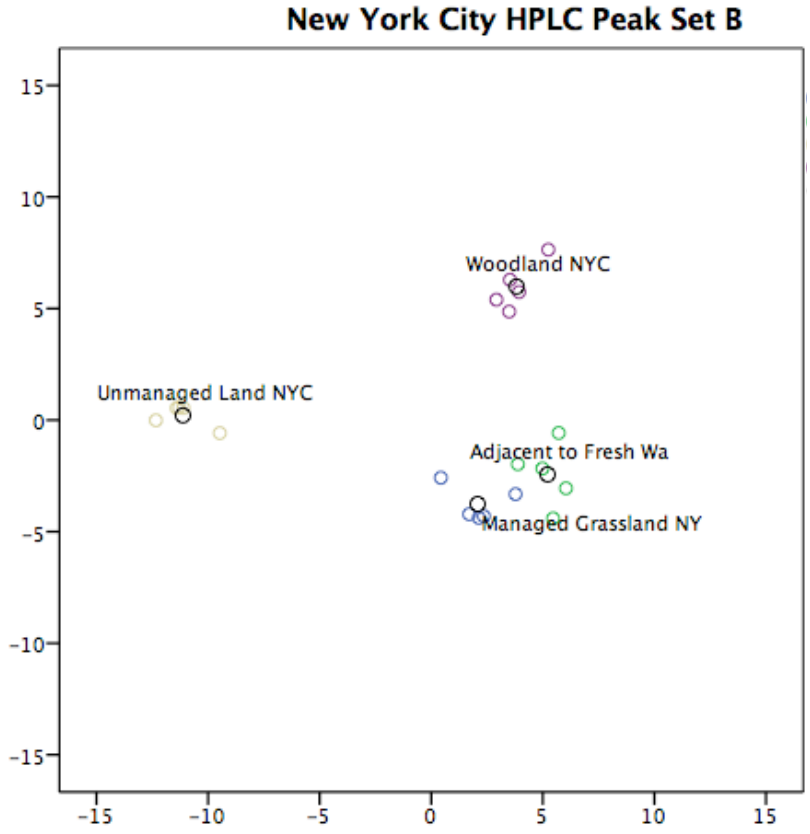


Figure 5.19 CDFA plots of sample scores for Function 2 vs. Function 1 for Peak Sets A and B for the London, Edinburgh, Aberdeen and New York City Sites. Group centroids are shown in black

5.4 Discussion

Systematic visual inspection of the anonymised chromatograms resulted in the correct grouping of all samples and it is encouraging that classification of such complex data was possible at such a high level of accuracy was achievable, particularly in a blind trial. As with other techniques which require detailed visual examination by an experienced analyst, the classification process was very time consuming. In this study it took around two hours to compare the chromatograms, which somewhat limits the practicality of this approach to data analysis for routine forensic analyses. It was not possible to visually identify any potential marker peaks on the basis of retention time and peak height, as the data was extremely complex and variable.

Performing CDFA on the entire data set produced excellent results, with 100% accuracy at the 99% confidence interval, however the data processing step was extremely labour intensive, due to the large number of peaks observed in the chromatographic data, and each peak requiring manual integration and classification for every analysis. The number of peaks in each chromatogram presented an enormous challenge in terms of ensuring that the data were correctly formatted for statistical analysis, since each of the 100+ peaks in one sample had to be compared to each of the 100+ peaks in the 19 other samples, and not all of the peaks were in every sample. It was difficult to ensure like-for-like comparisons were made due to natural, slight variations in the peak retention times from run to run, and the fact that many peaks were poorly resolved, and very close in retention time to their neighbours. As a result, the time taken to prepare the Chemstation data for analysis in SPSS was in the order of three weeks per set of 20 samples, which is an unacceptable time scale for urgent case work, and increases the labour cost associated with the analysis such that it is unlikely that it would be deemed useful by commercial or publically funded forensic science service providers.

This study has shown that measuring only these smaller sets of peaks, pre-selected for their discriminatory value, reduces the time taken to analyse the data from somewhere in the order of weeks, to less than one day. Peak sets A and B, both provided ways to distinguish the four close-proximity locations at each site based on visual assessment of the profiles, by peak ratios or the presence or absence of a particular peak. These results were consistent across the range of geographical locations used in the study, which indicates that the discriminatory ability of the technique is not affected by the underlying geology of a site, and the suitability of the technique for use at different sites and times of year is explored in greater detail in Chapter 6. Measurement of these specific peaks not only provides an accurate method of visually comparing locations within a site, but when combined with further processing of the

data using CDFA the ability of the technique to discriminate between these close proximity locations was excellent, demonstrating that removing the extra peaks from the analysis has minimal effect on the experimental outcome.

The improved data analysis method, using the two peak marker sets instead of the full dataset, significantly increases the potential impact of this HPLC method as a tool for the discrimination of geoforensic samples. The confirmation provided here of the suitability of these two peak sets for use as markers for geoforensic profiling allows the use of automated integration and peak identification in future studies, which could allow the data to be prepared in seconds by the chromatography software ready for CDFA analysis in SPSS, potentially reducing overall the data analysis time to a few minutes.

The use of automated data processing on these selected peaks would also improve the precision between the replicates at each sample location, as the software ensures greater consistency in the integration of each peak. Furthermore, with fewer peaks to quantify per sample, it is easier to accurately assign the peaks of interest and reduce misclassification errors resulting from coeluting peaks and poor chromatographic resolution.

There was noticeable variability in the data, evident from the error bars displayed in figures 5.11-5.18, which was expected due to the natural heterogeneity of soil as a sample. This variation would likely have been reduced had the five samples for each location been homogenised then sub-sampled, however since there was no pooling of the five replicates to offer an appropriate forensic context, the results are a better reflection of the true variability within each location. Homogenisation of samples must only be performed after careful consideration of specific case circumstances (13), since there may be small quantities of diagnostic or characteristic compounds, that are essential to the interpretation of the results, present in a discrete soil aggregate or individual sample point, and this information may be lost if sample mixing dilutes such compounds to below the limit of detection for the method. Understanding the degree of variability in the profiles at a location of forensic interest is essential when making comparisons of control and evidentiary samples, therefore preservation of intra-location variation is key to correctly interpreting geoforensic evidence.

It may however be possible, and indeed appropriate, to improve the precision between replicates by homogenising the individual replicate sub-samples through grinding and sieving each sub-sample prior to extraction. This could be achieved without the loss of any information necessary for the correct forensic interpretation, since any profile variability within the few-centimeter cross section of soil at each point in the sample grid is unlikely to be relevant to crime reconstruction. This approach would produce a sub-sample of uniform

content, which is likely to be more similar to the averaged result for other homogenised sub-samples at that location. Grinding and sieving would also improve the extractability of organic compounds in the soil, thereby increasing their concentration in sample solutions and the size of the corresponding peaks, which reduces the intra-location variability since the analytical precision of the HPLC method improves with peak size. In addition, through the investigation of more rigidly specified location types, in particular by exercising greater control over the variation of surface vegetation at each location type when selecting sample locations. This approach may also lead to the identification of location-specific markers or profiles that are indicative of particular, well defined land use or vegetation types, and thereby be valuable for the 'seek and find' investigations in addition to the comparison between exhibits.

The classification accuracy rate achieved for the two sets of marker peaks was very high in both cases, with both sets achieving 100% accuracy at three out of four sites. The results were slightly better for peak set A, which misclassified only one sample across the whole study, compared to peak set B which misclassified two samples. In this regard, set A can be said to offer superior results, however the data for peak set B had the advantage of being much more easily interpreted by visual examination of the profiles, due to having fewer variables to compare. With fewer peaks to identify, classify and analyse, set B also offers the added benefit of reduced data analysis time compared to set A.

In order to develop this method for use in forensic casework, it will be necessary to validate the ability of this method to reliably detect and quantify the peaks of interest, which requires further research to identify the compounds and isolate purified extracts for use as standards in the method validation. Characterisation of the smaller set of compounds in set B could be significantly quicker and optimising the chromatography of fewer peaks would make validation far easier. However given their low concentration it would take longer and be more costly to extract sufficient quantities for use as standards. Furthermore it may be necessary to introduce a pre-concentration step to the sample preparation method in order to achieve robust analysis using peak set B.

5.5 Conclusions

This study has shown it is not necessary to use all the peaks present in the HPLC profile of a sample in the data analysis and sample comparison steps. Visual comparison of the initial chromatograms showed that there were regions where the chromatography was noticeably different between groups, and by focusing on these three regions only it was possible to classify the samples with 100% accuracy. Although this visual assessment was naturally

subjective, it was possible to select markers in a more objective manner through the use of the "Subselect" analysis in R and the informed choice of large, easily identifiable and reliably quantifiable peaks. Near 100% classification accuracy rates were achieved at all four sites in each of the four locations tested, using these two sets of markers. This research provides investigators with two effective data analysis strategies to use in future geoforensic studies, allowing the data analysis method to be chosen to suit the priorities of the case specific details. Where sample amounts are limited the use of the larger peak set will be valuable, and where quicker analysis is required the smaller peak set offers timely analysis. The results of this research show that the newly developed HPLC and data analysis methods provide significant scope for highly discriminatory, routine analyses to be performed on geoforensic samples from close-proximity sites.

6 Geographic and Seasonal Variability

6.1 Introduction

Having established that the chromatograms at a particular location are suitably complex to allow accurate discrimination at a forensically relevant spatial scale in the London site (see Chapter 4) and having further developed the sample preparation and data analysis methods (see Chapters 3 and 5) such that the technique was practical for routine use, it was necessary to ascertain whether these results could be replicated across sites with a range of underlying geologies, in order to establish whether the test is valid outside of London, and at different times of year, to test the robustness of the technique to seasonal changes in soil composition.

6.2 Methodology

Samples were collected from the four sites described in Chapter 5, in London, Edinburgh, Aberdeen and New York City, at regular intervals over an eighteen month period and the sampling plan is summarised in Table 6.1, ensured that one set of results allowed comparison of all four cities, three sites were sampled for a full year, and one of the cities was sampled over all six time points, allowing year-on-year comparisons to be made. Samples were initially collected in London in June 2013, then in again in November 2013. All four cities were sampled in January 2014, and the three UK sites were sampled again in April, June and November 2014.

Table 6.1: Summary of Sites and Time points sampled.

	June 2013	November 2013	January 2014	April 2014	June 2014	November 2014
Brockwell Park, London	x	x	x	x	x	x
Lochend Park, Edinburgh			x	x	x	x
Craigiebuckler, Aberdeen			x	x	x	x
Central Park, New York City			x			

Samples were collected and stored as described in Chapter 3 and analysed using the final analysis method described in Chapter 5. All peaks in the resulting chromatograms were

integrated using Chemstation software and then processed in Excel to remove peaks below the LOQ.

The two HPLC peak subsets discussed in Chapter 5 both gave good accuracy in sample discrimination, as presented in section 5.3, and each offered benefits over the other in certain case circumstances, as the peaks in set A gave slightly better accuracy rates had larger responses and could therefore be detected in smaller amounts of soil in cases where only small traces of sample were available, while the smaller number of markers in peak set B offered easier and quicker data analysis, in cases where speedy analysis is required, with the additional benefit of potentially making the validation of the technique much easier, quicker and cost effective. As a result of these varied benefits, both peak sets were used in the analysis presented in this chapter. The relevant peaks from each peak set were extracted from the Excel data sets and CDFA was performed using SPSS.

6.3 Results

The accuracy and significance of the CDFA for each site and time point are summarised in Table 6.2 for peak set A and Table 6.3 for peak set B. Further discussion of these results is provided for each individual analysis, in sections 6.3.1, 6.3.2, 6.3.3, and 6.3.4, for the Edinburgh, Aberdeen, London and New York City sites, respectively.

The results for the Winter (January 2014) time point have also been presented in Chapter 5 to discuss the respective merits of peak sets A and B, and will also be discussed in Chapter 7 in comparison to the results obtained for these samples using more established geoforensic analytical techniques. The data for the Winter samples are therefore presented in this chapter only to examine the effects of changing the geographic position of the sampling site, and the time of year that the samples are collected on the ability to reliably discriminate the samples and any implications these may have on how the technique could be used in practice.

Table 6.2 Canonical Discriminant Function Results for Peak Set A

HPLC Profiles	Classification Accuracy %	Wilks' Lambda Significance test of functions			% Variance Function 1	% Variance Function 2	% Variance Function 3
London		1-3	2-3	3			
Summer 2013	100.0	0.000	0.000	0.007	91.7	7.4	1.0
Autumn 2013	89.5	0.009	0.303	0.882	80.0	18.0	2.0
Winter 2014	100.0	0.000	0.002	0.034	89.7	7.0	3.3

Spring 2014	100.0	0.000	0.028	0.361	81.2	15.8	3.0
Summer 2014	100.0	0.000	0.035	0.273	83.9	96.7	100.0
Autumn 2014	100.0	0.000	0.070	0.536	91.4	7.4	1.1
Edinburgh							
Winter 2014	100.0	0.000	0.000	0.022	88.9	8.3	2.7
Spring 2014	100.0	0.045	0.149	N/A*	72.1	27.9	N/A*
Summer 2014	100.0	0.000	0.019	0.512	90.9	8.5	0.6
Autumn 2014	100.0	0.000	0.027	N/A*	90.9	9.1	N/A*
Aberdeen							
Winter 2014	94.7	0.001	0.147	0.531	90.6	7.8	1.5
Spring 2014	100.0	0.000	0.000	0.003	45.0	34.6	20.4
Summer 2014	100.0	0.000	0002	0.091	96.8	2.7	0.5
Autumn 2014	100.0	0.000	0.009	0.090	90.5	7.0	2.5
New York							
Winter 2014	100.0	0.000	0.000	0.005	92.3	5.8	1.9

* Due to the porosity of the soil, no chromatograms could be obtained for soils Adjacent to Fresh Water in Spring or Autumn therefore only two canonical functions were generated, significance values are for functions 1 and 2, and function 2, respectively.

Table 6.3 Canonical Discriminant Function Results for Peak Set B

HPLC Profiles	Classification Accuracy %	Wilks' Lambda Significance test of functions			% Variance Function 1	% Variance Function 2	% Variance Function 3
London		1-3	2-3	3			
Summer 2013	100.0	0.000	0.328	0.711	95.2	3.8	0.9
Autumn 2013	80.0	0.377	0.729	0.841	67.0	24.7	8.4
Winter 2014	90.0	0.000	0.041	0.684	84.8	13.9	1.3
Spring 2014	100.0	0.000	0.052	0.929	86.4	13.1	0.5
Summer 2014	100.0	0.000	0.005	0.119	87.4	9.8	2.8
Autumn 2014	95.0	0.062	0.526	0.418	78.6	11.3	10.1
Edinburgh							

Winter 2014	100.0	0.000	0.018	0.397	62.4	32.5	5.0
Spring 2014	100.0	0.021*	0.998*	N/A*	72.9	27.1	N/A*
Summer 2014	94.7	0.000	0.000	0.230	53.2	45.0	1.8
Autumn 2014	100.0	0.001*	0.353*	N/A*	97.4	2.6	N/A*
Aberdeen							
Winter 2014	100.0	0.000	0.000	0.014	97.4	2.4	0.2
Spring 2014	100.0	0.000	0.002	0.161	74.9	21.7	3.5
Summer 2014	100.0	0.000	0.000	0.004	95.4	4.1	0.5
Autumn 2014	95.0	0.145	0.378	0.440	60.7	24.6	14.7
New York							
Winter 2014	100.0	0.000	0.000	0.071	73.3	24.1	2.5

6.3.1 Edinburgh

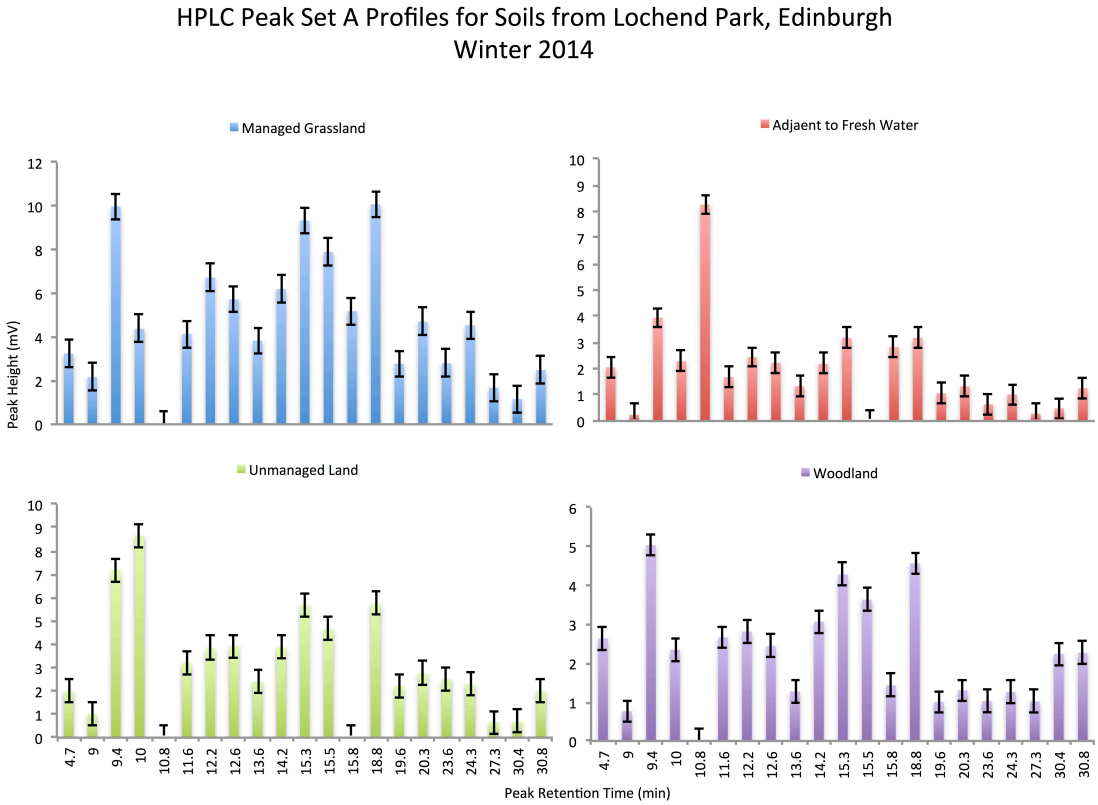
6.3.1.1 HPLC Profiles for Peak Set A

The results of the visual comparison of the Peak Set A profiles obtained at Lochend Park, Edinburgh and the results of the CDFA using the peaks in set A as variables, are summarised below for the four 2014 time points detailed in Table 6.1.

Winter 2014

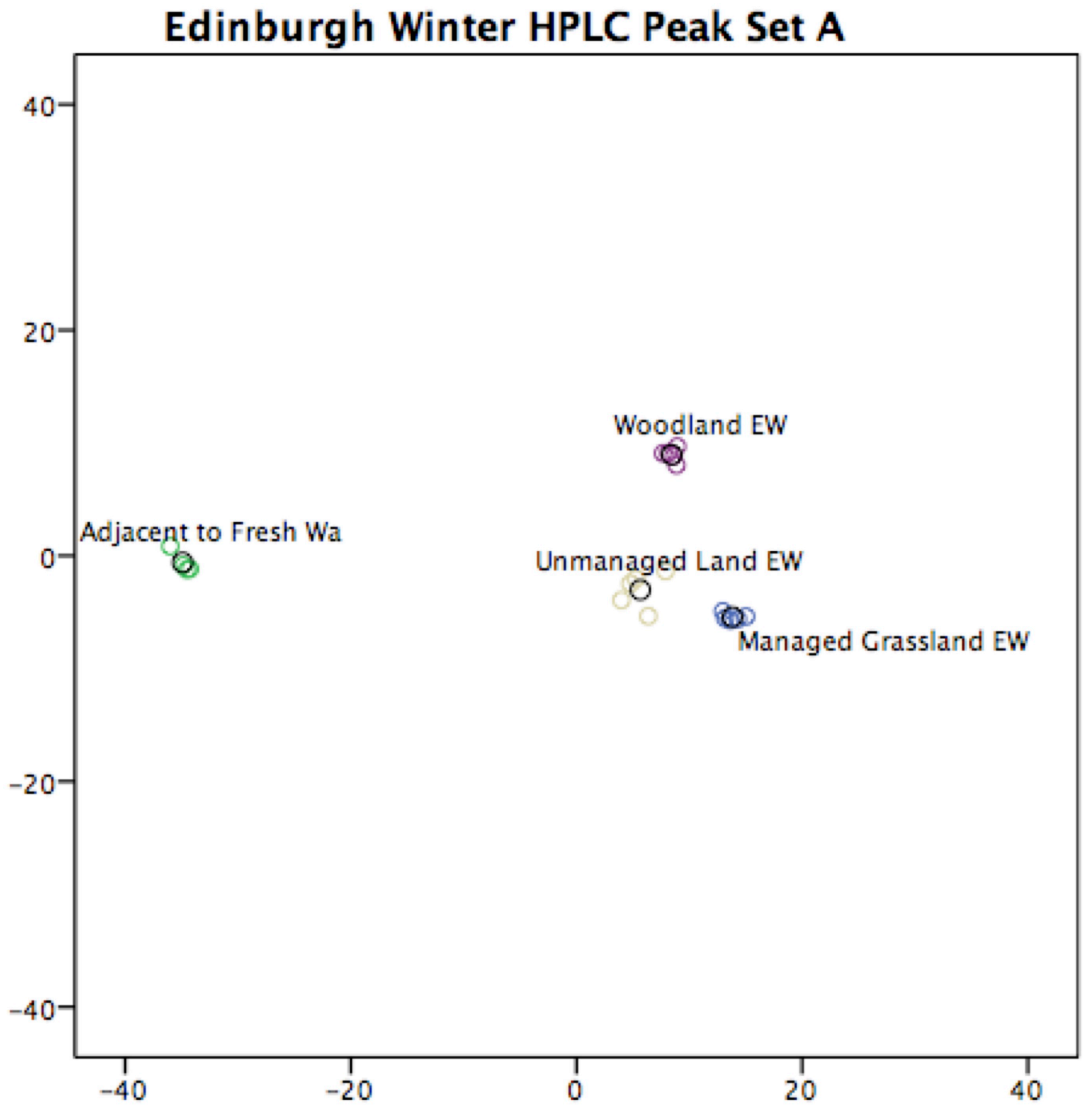
Figure 6.1 presents the HPLC profiles for peak set A. As discussed in Chapter 5, all four locations within Lochend Park, Edinburgh could be discriminated on the basis of the profiles of HPLC peak set A. The peak at 10.8min was indicative of soils adjacent to fresh water, since this was absent at the other three locations, while soils from unmanaged land were the only samples that did not contain a peak at 15.8min. Woodland and managed grassland samples were very similar at this site, however at the managed grassland location, the ratio of the peak height at 9min relative to the peak at 4.7min was much larger, at 3.5:1, than for woodland samples, where this ratio was 1.5:1. In addition, the peaks at 20.3 and 24.3min were 69% and 63% larger for managed grassland than the peaks at 19.6 and 23.6min, respectively, but for woodland soils these peaks were only 30% and 26% larger. For woodland soils the peak at 30.4min was approximately equal in size to the peak at 30.8min, while for managed grassland the peak at 30.4min was half the size of the peak at 30.8min.

Figure 6.1 Seasonal Changes to Peak Set A Profiles- Edinburgh, Winter.



The statistical analysis using peak set A produced three canonical discriminant functions (Figure 6.2, Table 6.2) which explained 88.9%, 8.3% and 2.7% of the variance in the samples, respectively, and gave 100% accuracy in grouping the samples using these functions, and this discrimination was significant at the 99% confidence interval (p= 0.000) when all three functions were used together. The separation achieved using the first two canonical functions is displayed in the scatter plot in Figure 6.2.

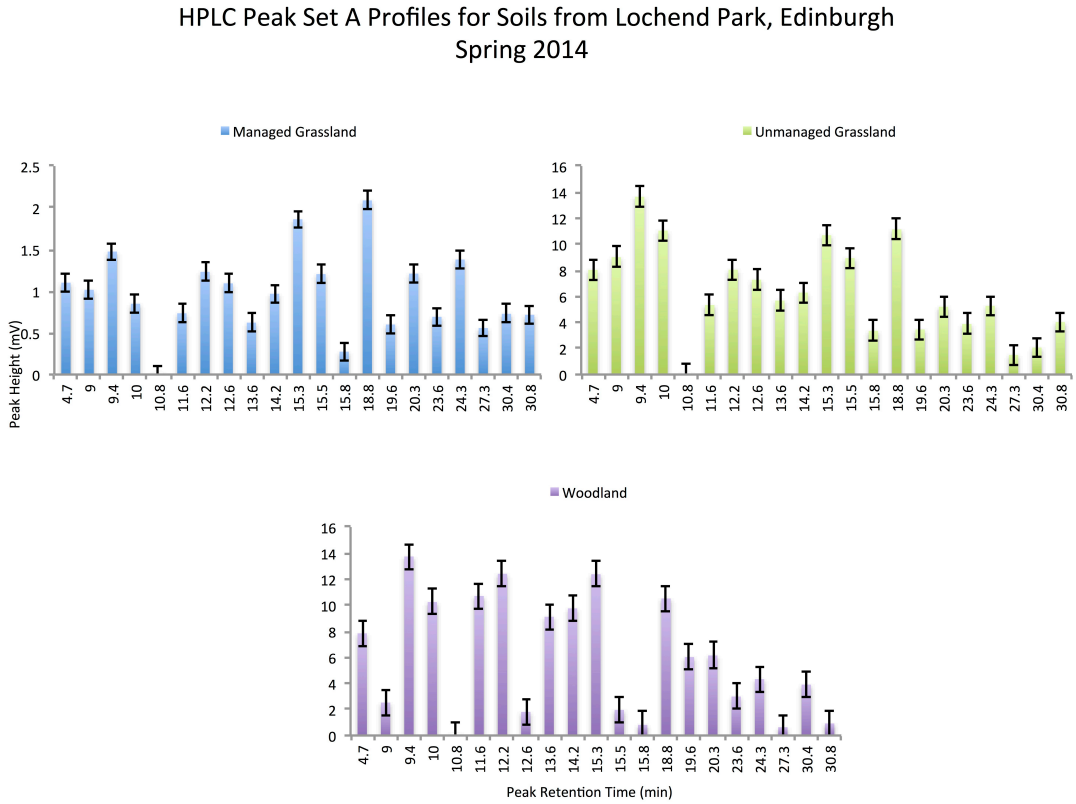
Figure 6.2 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set A- Edinburgh, Winter.



Spring 2014

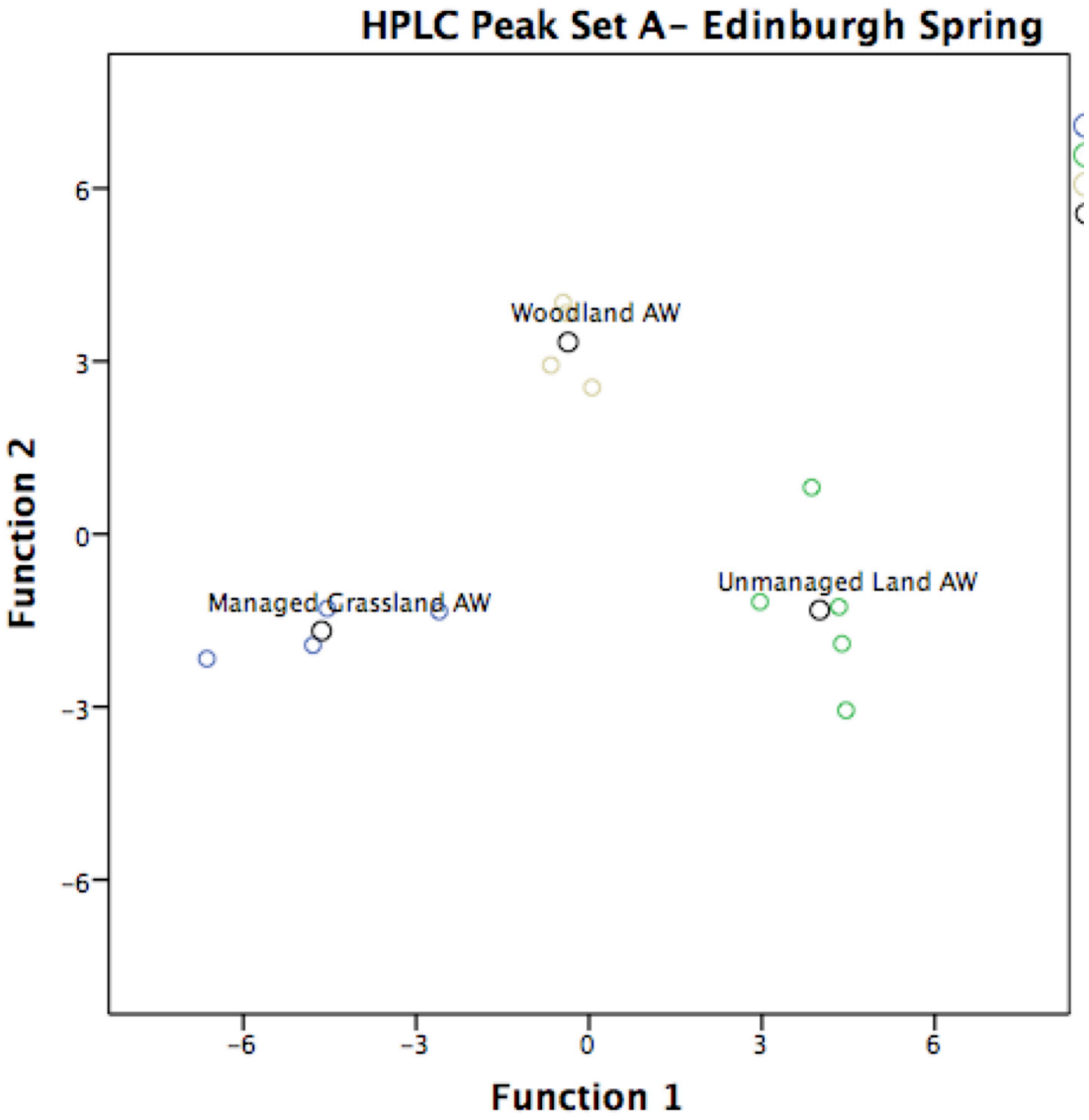
Figure 6.3 presents the results of the HPLC analysis for peak set A. The soil at the location adjacent to water for the Edinburgh site was very porous at the Spring time-point and therefore was so absorbent that it was not possible to extract the supernatant sample liquid for injection onto the HPLC, therefore the sample set contained only three location groups. Nevertheless, the profiles for HPLC peak set A at these three locations could be discriminated visually. The woodland soils could be discriminated from the other two locations by the ratio of the peak at 9.4min to the peak at 9min, which was 5.4:1 compared to 1.5:1 for managed grassland and unmanaged. The profiles for peak set A for managed grassland could be distinguished from those for unmanaged land by the ratio of the peak at 15.5min to the peak at 15.8min, which was 7.7:1 for the former and 3.3:1 for the latter location.

Figure 6.3 Seasonal Changes to Peak Set A Profiles- Edinburgh, Spring



The CDFA (Figure 6.4, Table 6. 2) produced two functions which explained 72.1% and 27.9% of the variation between the groups, respectively. The analysis correctly grouped the samples with 100% accuracy and the discrimination between the groups was statistically significant at the 95% confidence interval (p=0.045).

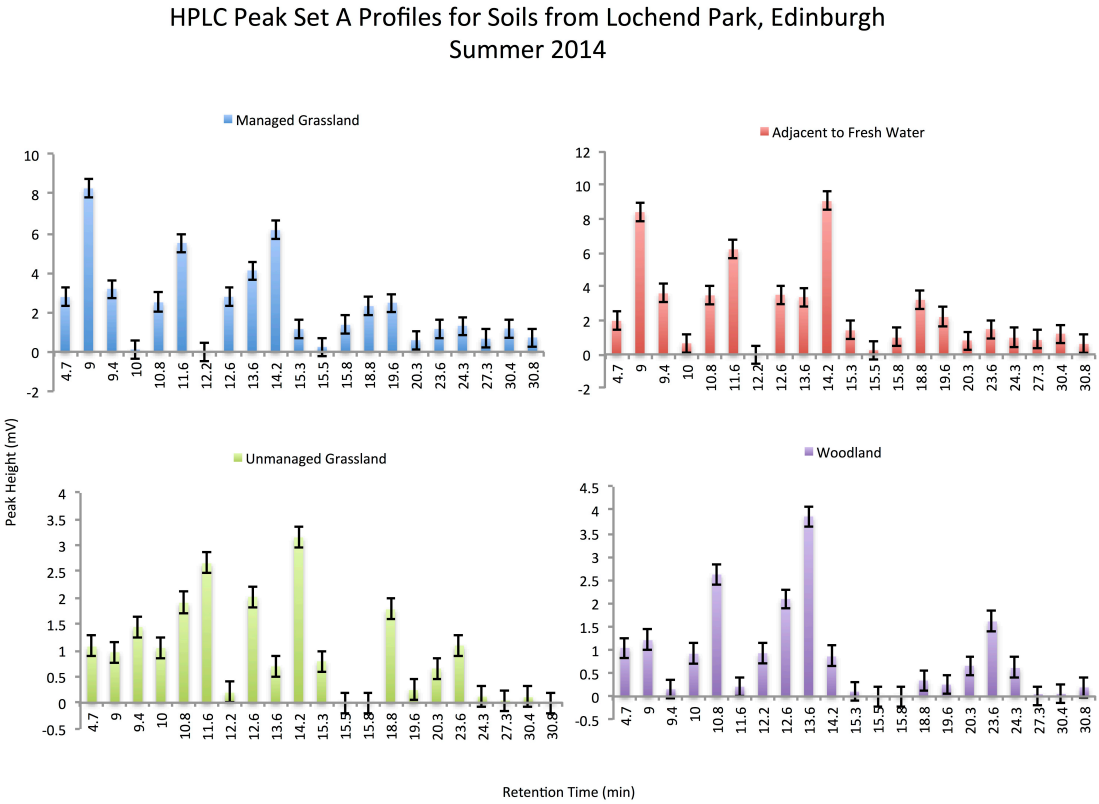
Figure 6.4 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set A- Edinburgh, Spring.



Summer 2014

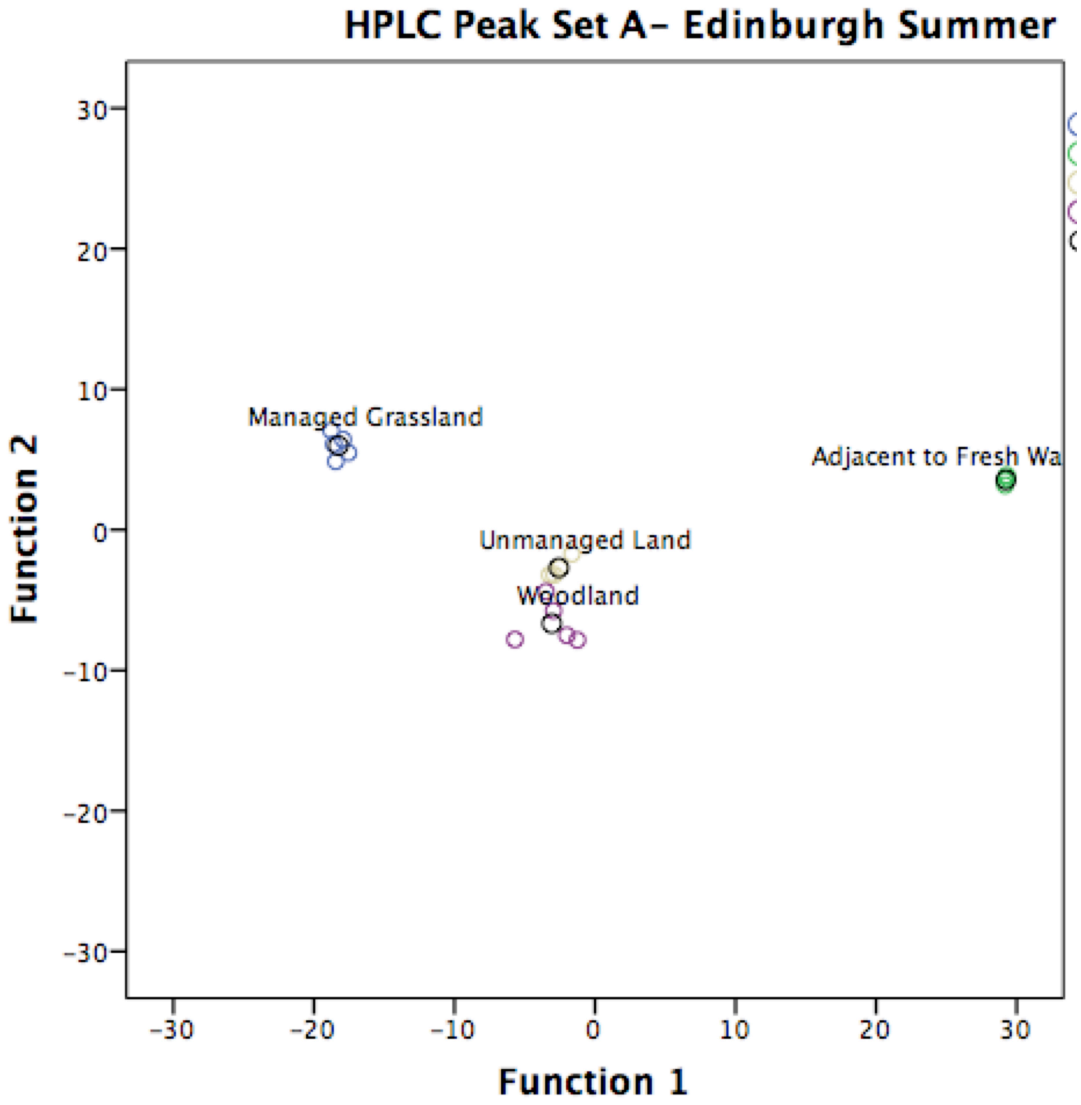
In Summer, the profiles for peak set A (Figure 6.5) in soils from unmanaged land and woodland could be distinguished from those adjacent to fresh water and those from managed grassland by the ratio of the peaks at 9min and 4.7min, which were 1:1 (1.1:1 and 0.9:1, respectively) for the former pair and 4.2:1 and 3:1 for the latter pair, respectively. Unmanaged land and woodland soils could be distinguished from one another since the largest peak in the former was at 14.2min at while the largest peak in woodland soil was at 13.6min, and soils adjacent to fresh water could be discriminated from managed grassland since the ratio between 13.6 and 12.6 was 1:1 for the former, and 1.5:1 for the latter.

Figure 6.5 Seasonal Changes to Peak Set A Profiles- Edinburgh, Summer



The statistical analysis (Figure 6.6, Table 6.2) produced three canonical discriminant functions which explained 90.0%, 8.5% and 0.6% of the variance in the samples, respectively and the samples were grouped with 100% accuracy using these three functions. Furthermore, the discrimination using these profiles was statistically significant at the 99% confidence interval (p=0.000)

Figure 6.6 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set A- Edinburgh, Summer.

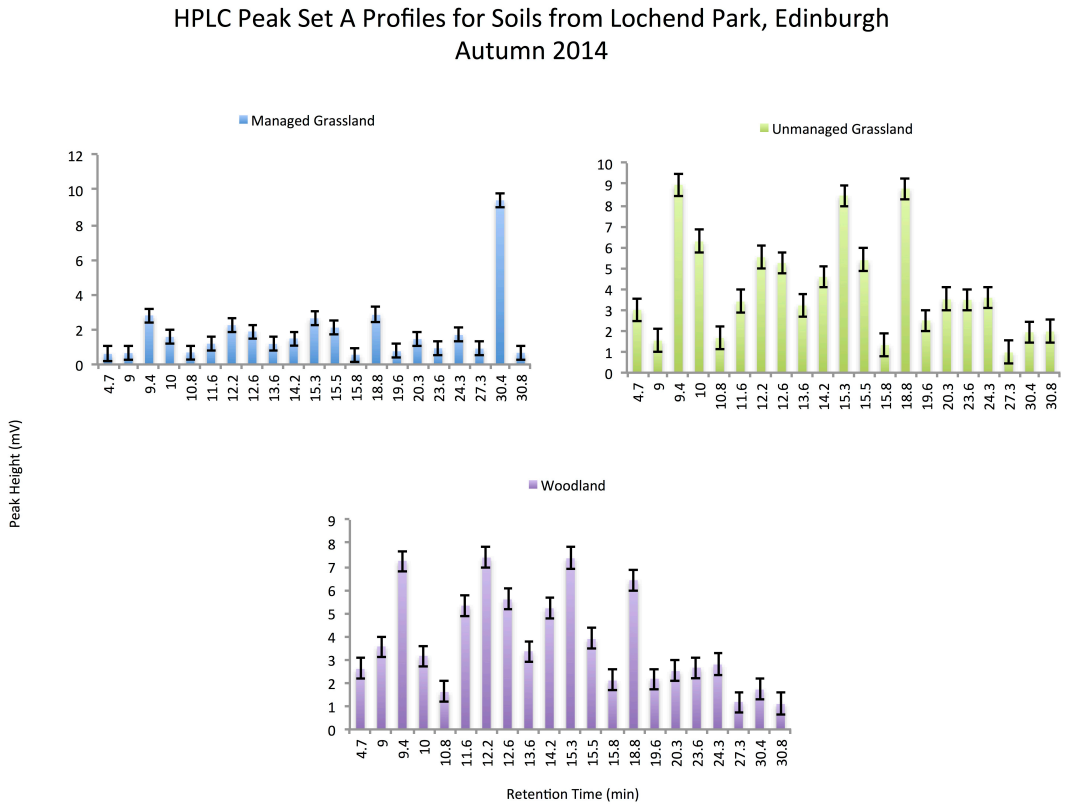


Autumn 2014

The soils at the Autumn time point at the location adjacent to fresh water in Edinburgh were again highly porous, as was the case at the spring time point, which meant that the acetonitrile was retained within the pores in the soil after sample preparation and chromatograms could not be obtained for this location. The profiles for peak set A (Figure 6.7), however, allowed visual discrimination of the three remaining locations. Managed grassland was easily identified by the very large peak at 30.4min which was over three times larger, at 9.40mV, than all other peaks, which ranged from 0.55mV to 2.84mV. The relative size of the peaks at 9min to the peaks at 9.4min allowed the soils from woodland and unmanaged land to be discriminated from one another, for woodland the peaks were 3.57mV and 7.24mV

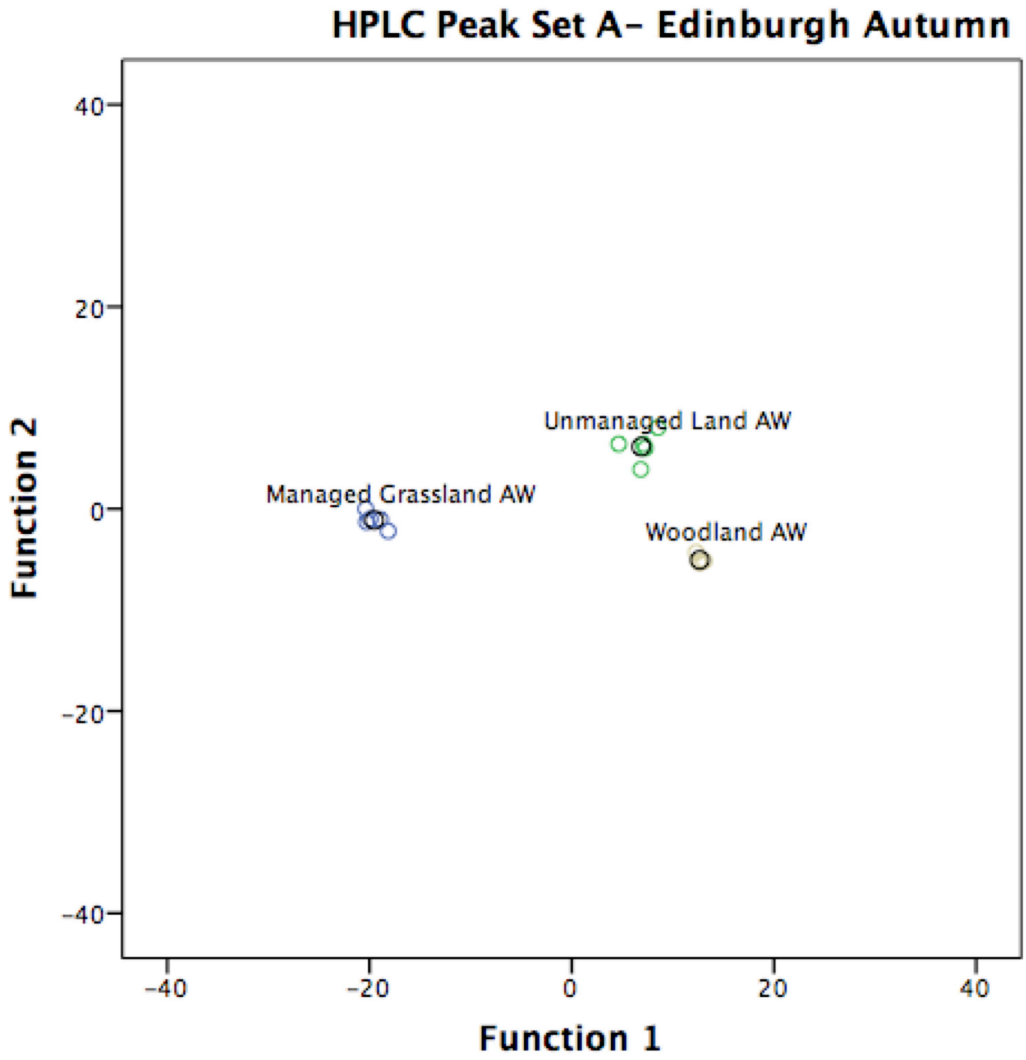
respectively, while for unmanaged land the same peaks were 1.53mV and 8.97mV meaning that the ratio was three times larger, at 6:1, for the unmanaged land than it was for woodland, where the ratio was 2:1.

Figure 6.7 Seasonal Changes to Peak Set A Profiles- Edinburgh, Autumn



For these profiles, the CDFA (Figure 6.8, Table 6.2) again produced only two functions, to separate the three groups, which explained 90.9% and 9.1% of the variance in the samples. The functions allowed the samples to be discriminated with 100% accuracy and the discrimination was significant at the 99% confidence interval (p=0.000)

Figure 6.8 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set A- Edinburgh, Autumn.



6.3.1.2 HPLC Profiles for Peak Set B

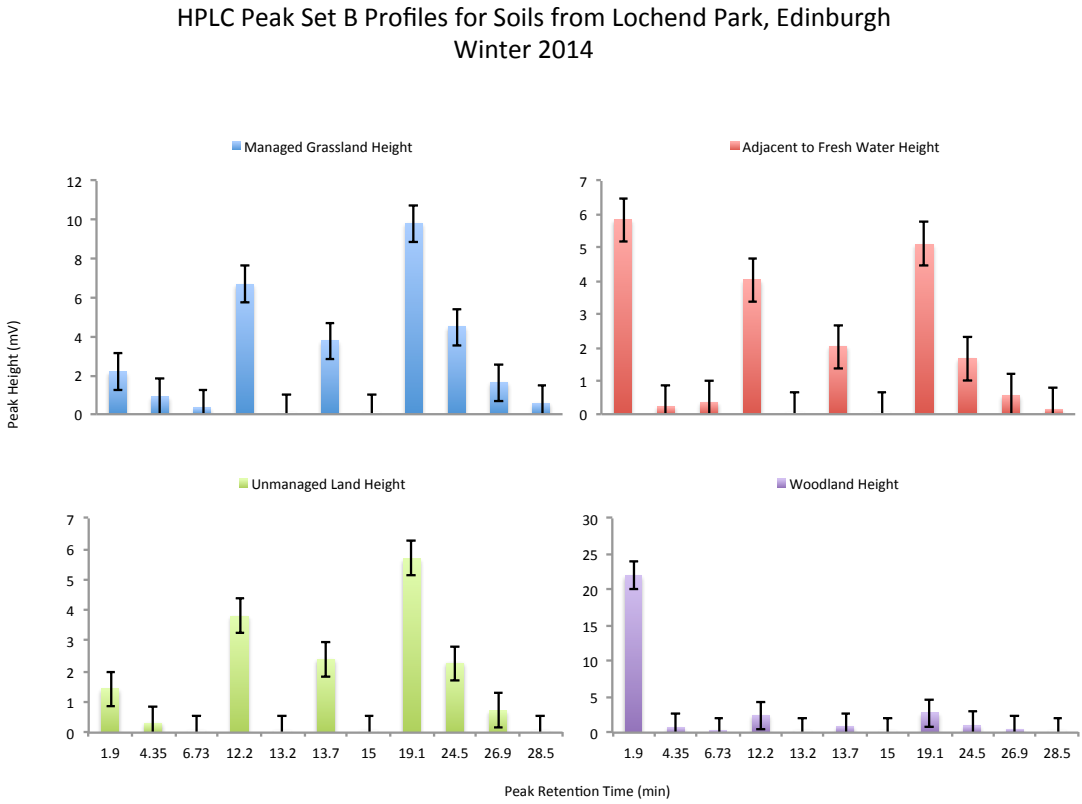
The Peak Set B profiles obtained at Lochend Park, Edinburgh are displayed below for visual comparison of the time points detailed in Table 6.1. alongside the results of the CDFA using the peaks in set B as variables.

Winter 2014

Figure 6.9 presents the profiles for HPLC peak set B, which are also presented in Chapter 5, and shows that the profiles were less easily discriminated at the four locations in Lochend Park, Edinburgh, however, each profile was visibly different from the others. The woodland samples could be separated from the other locations in that they had large peaks at 1.9min, which, at 22.0mV, were approximately ten times the size of the next largest peaks at 12.2 and

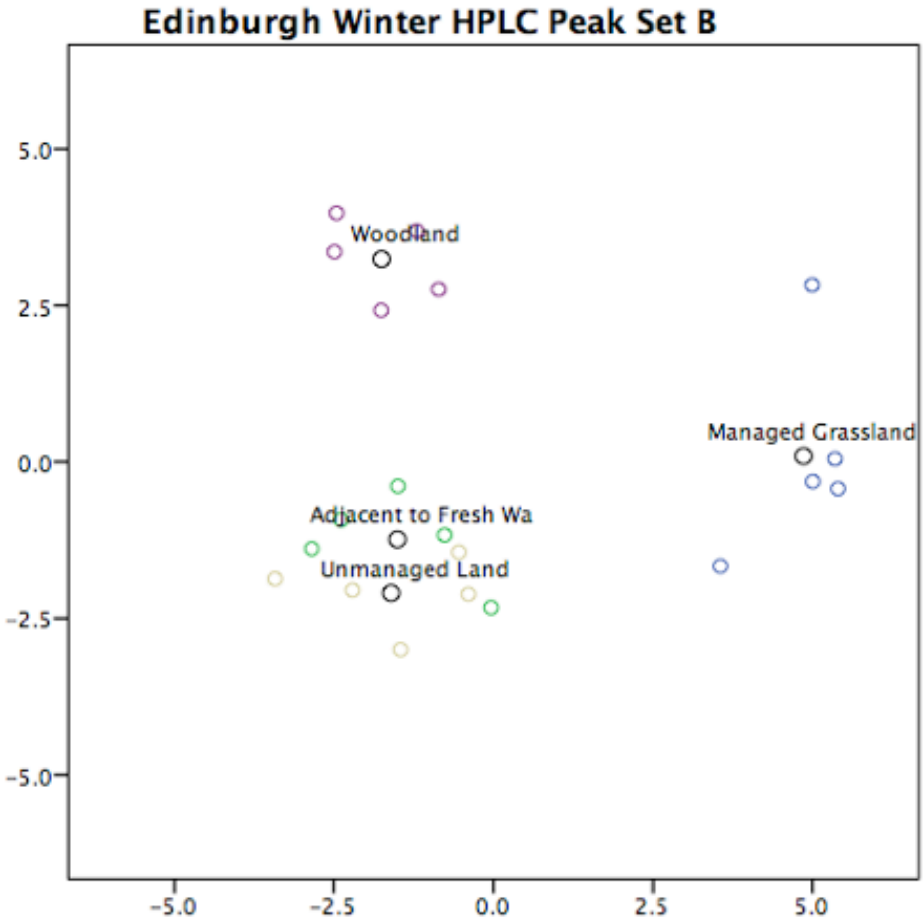
19.1min at 2.4 and 2.8mV, respectively. The distinguishing feature for soils adjacent to fresh water was that the two largest peaks at 1.9 and 19.1min, and these were also very similar in size to one another. The largest peak in the profiles for managed grassland was 9.8mV, which was approximately twice the size of the same peak in the latter, at 5,6mV, and the small peaks present in the managed grassland profiles at 6.7 and 28.5min were absent in the samples from unmanaged land, allowing these two groups to be separated. However the relative uncertainty in measurements of peaks at this low level is far greater and therefore there can be much less certainty in the visual discrimination of these two locations using these low level peaks.

Figure 6.9 Seasonal Changes to Peak Set B Profiles- Edinburgh, Winter



For these profiles, the CDFA produced three functions, to separate the four groups, which explained 62.4%, 32.5% and 5.0% of the variance in the samples (Figure 6.10, Table 6.3). The functions allowed the samples to be discriminated with 100% accuracy and the discrimination was significant at the 99% confidence interval (p=0.000) when all three functions were used.

Figure 6.10 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set B- Edinburgh, Winter.

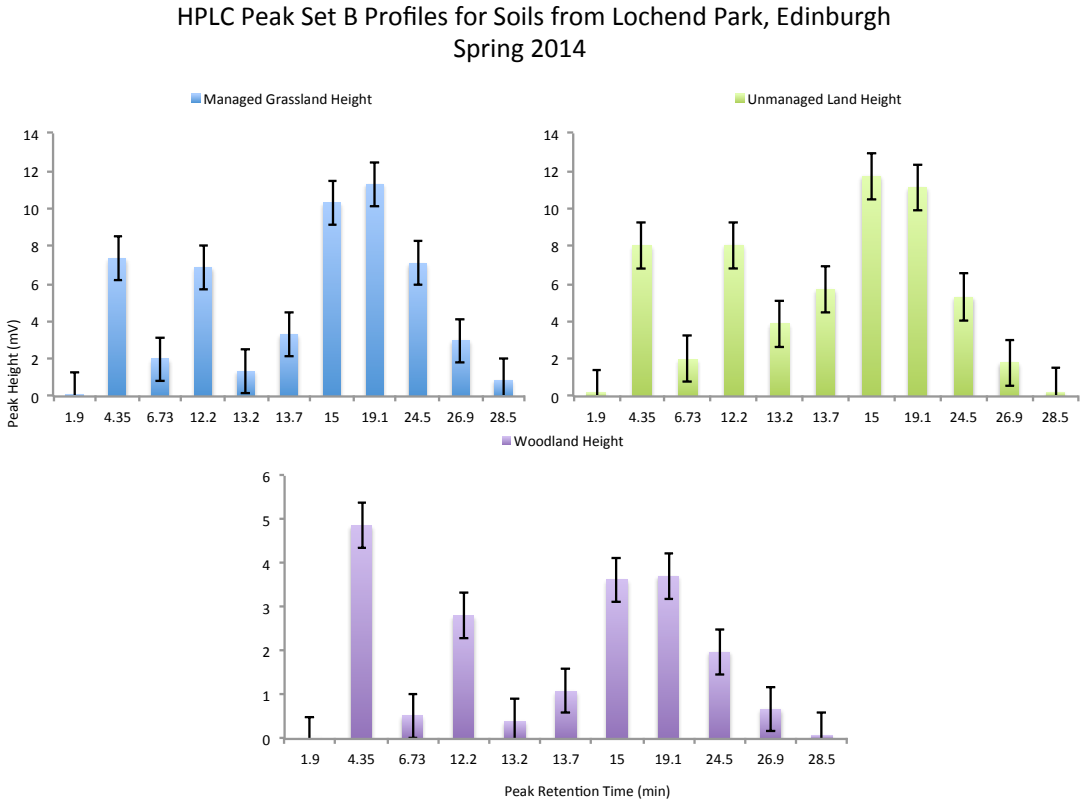


Spring 2014

As discussed for peak set A, insufficient sample aliquots could be extracted from the physical samples, therefore no chromatograms were obtained for the location adjacent to fresh water at this time point. The relative heights of the peaks at 4.35 and 19.1min allowed the woodland soils to be discriminated from the managed and unmanaged locations for the Edinburgh site in Spring (Figure 6.11). In woodland soils the peak at 4.35min was 33% larger than the peak at 19.1min and the ratio of the former to the latter was 0.76:1, while the peak at 19.1min was 28% and 35% smaller than the peak at 4.35min, for unmanaged land and managed grassland, with ratios of 1.4:1 and 1.5:1, respectively. To visually separate the managed and unmanaged locations, comparison of the relative sixes of the peaks at 13.7min and 24.5min proved useful, since the two peaks were of roughly equal size for unmanaged land, with a ratio of the latter to the former at 0.92:1, while for managed grassland the peak at

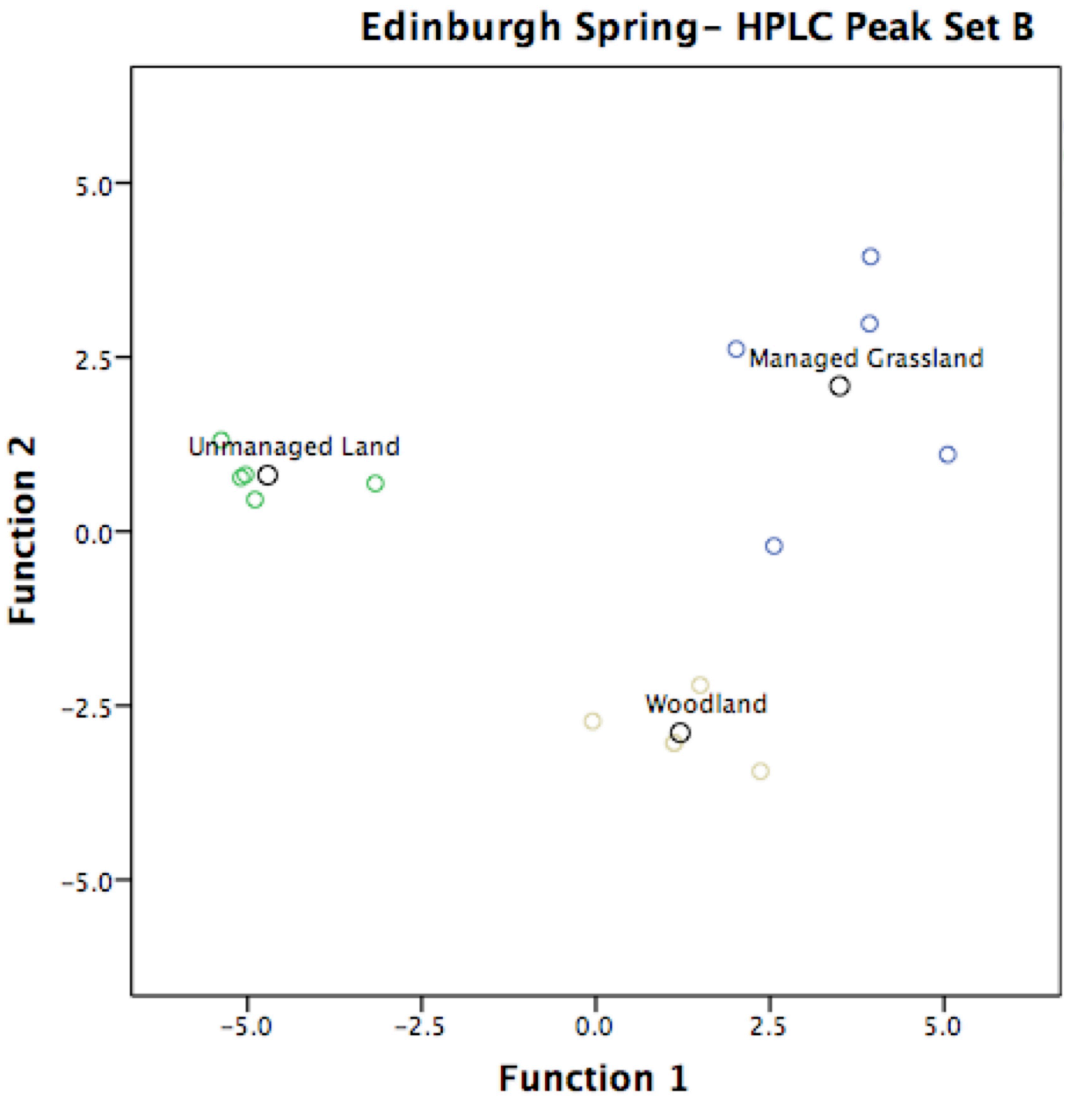
13.7min was 54% smaller than the peak at 24.5min, with a ratio of the latter to the former of 2.16:1.

Figure 6.11 Seasonal Changes to Peak Set B Profiles- Edinburgh, Spring



The CDFA (Figure 6.12, Table 6.3) grouped the Spring samples from Edinburgh with 100% accuracy, and the discrimination was statistically significant at the 95% confidence level ($p=0.021$) using the two canonical functions produced to separate the three sample groups, which accounted for 72.9% and 27.1% of the variance in the data.

Figure 6.12 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set B- Edinburgh, Spring.

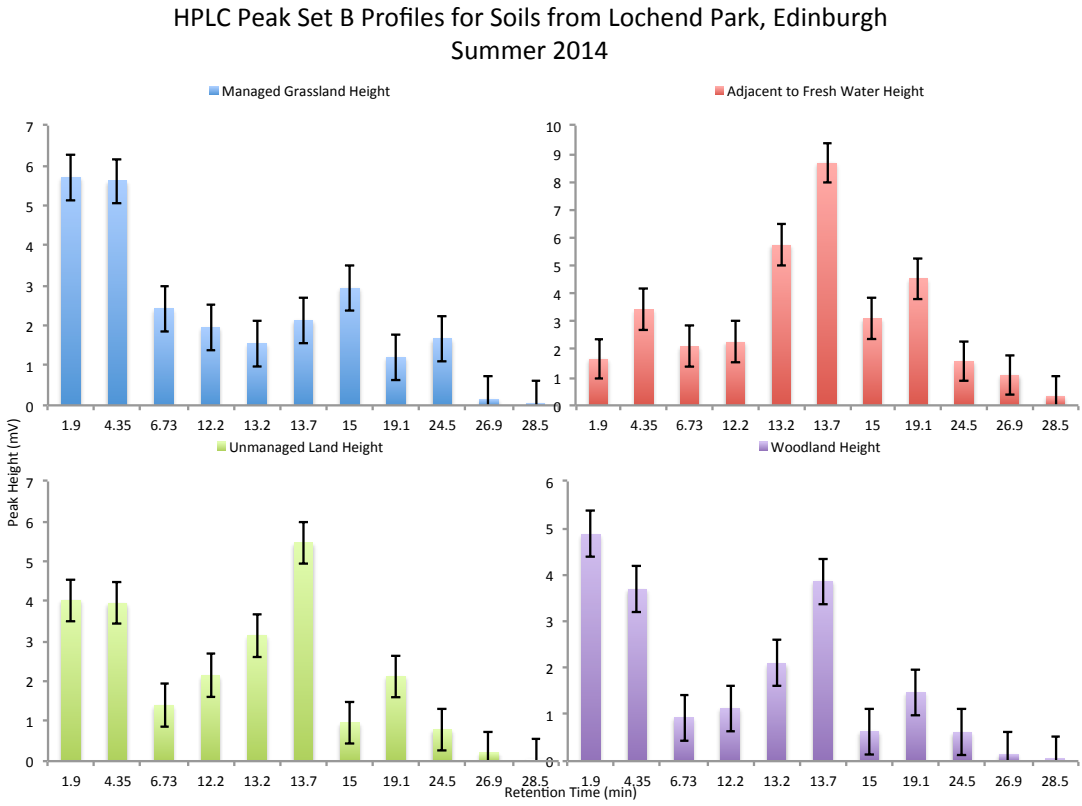


Summer 2014

The four location types in Lochend Park could be readily separated by visual comparison of the profiles obtained for peak set B (Figure 6.13). Ranking the peaks by size order was very useful for comparing the samples in this data set, managed grassland was characterised by having the two largest peaks at 1.9min and 4.35min, which were approximately equal in size, at 5.69 and 5.62mV and were 94% and 91%, respectively, larger than the next largest peak at 15min at 2.94mV. The unmanaged land also had two large, approximately equally sized peaks at 1.9min and 4.35min, however, at 4.02mV and 3.95mV respectively, these were ranked second

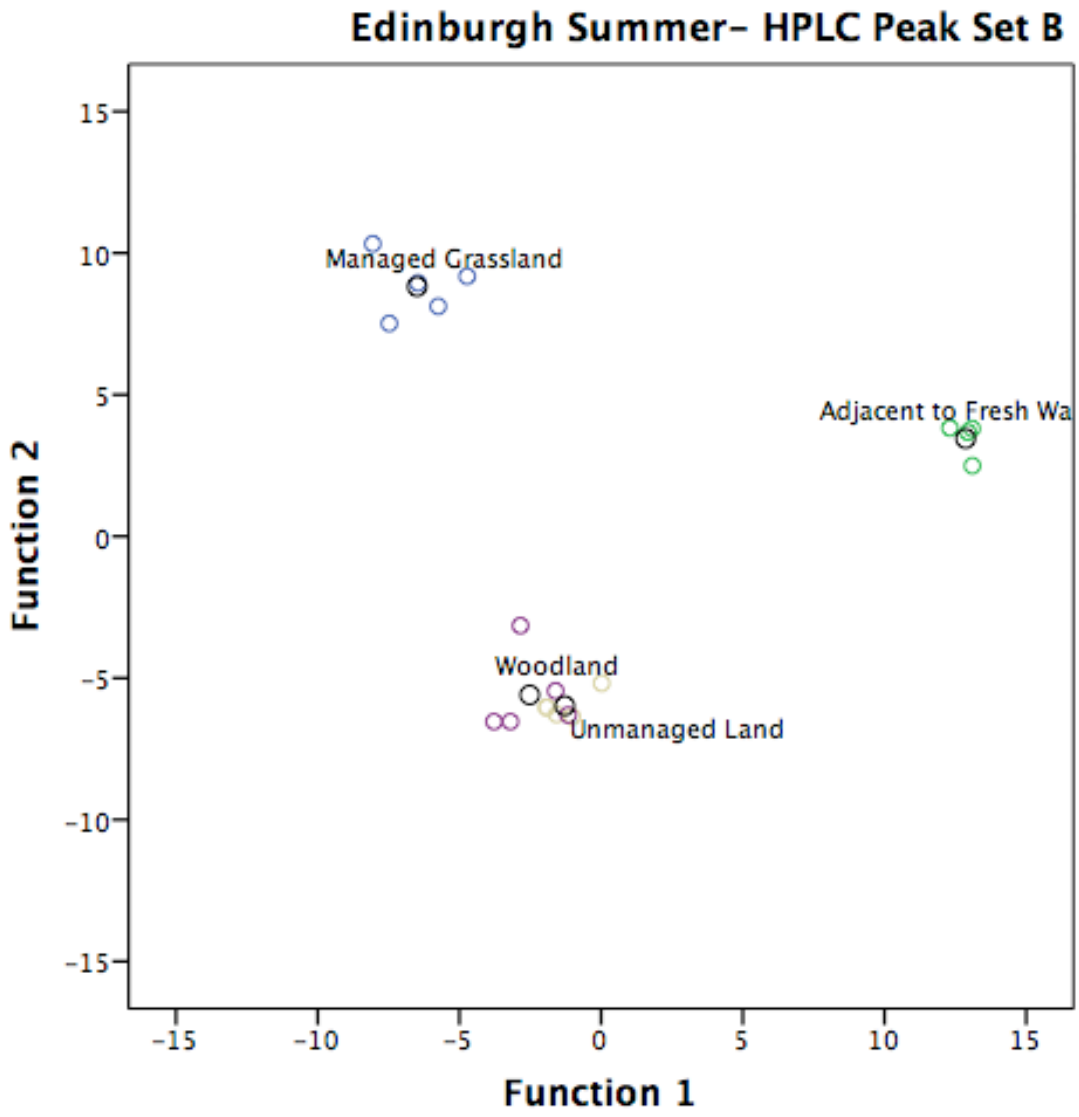
and third after the largest peak, at 13.7min, which was 36% and 38% larger, respectively, at 5.46mV. At the woodland location, the peaks at 1.9min, 4.35min , and 13.7min were also the three largest peaks, however they were ranked differently from the unmanaged land and were first, third and second largest, respectively and the peak at 1.9min was 32% larger than the peak at 4.35min, at 4.86mV and 3.69mV respectively, not equally sized as was the case for the unmanaged land. Unlike the other locations, the peaks at 1.9min and 4.35min were not major components of the profiles for soils adjacent to fresh water. Here, the largest peak was at 13.7min and this peak was 51% and 91% larger than the second and third largest peaks at 13.2min and 19.1min, respectively.

Figure 6.13 Seasonal Changes to Peak Set B Profiles- Edinburgh, Summer



In the CDFA (Figure 6.14, Table 6.3) for Edinburgh at the Summer time point, one woodland sample was misclassified as unmanaged land and the accuracy of grouping was 94.7%. The analysis resulted in a discrimination that was 100% accurate and statistically significant at the 99% confidence level (p=0.000). The three discriminant functions explained 62.4%, 32.5%, and 5.0% of the variability between the groups.

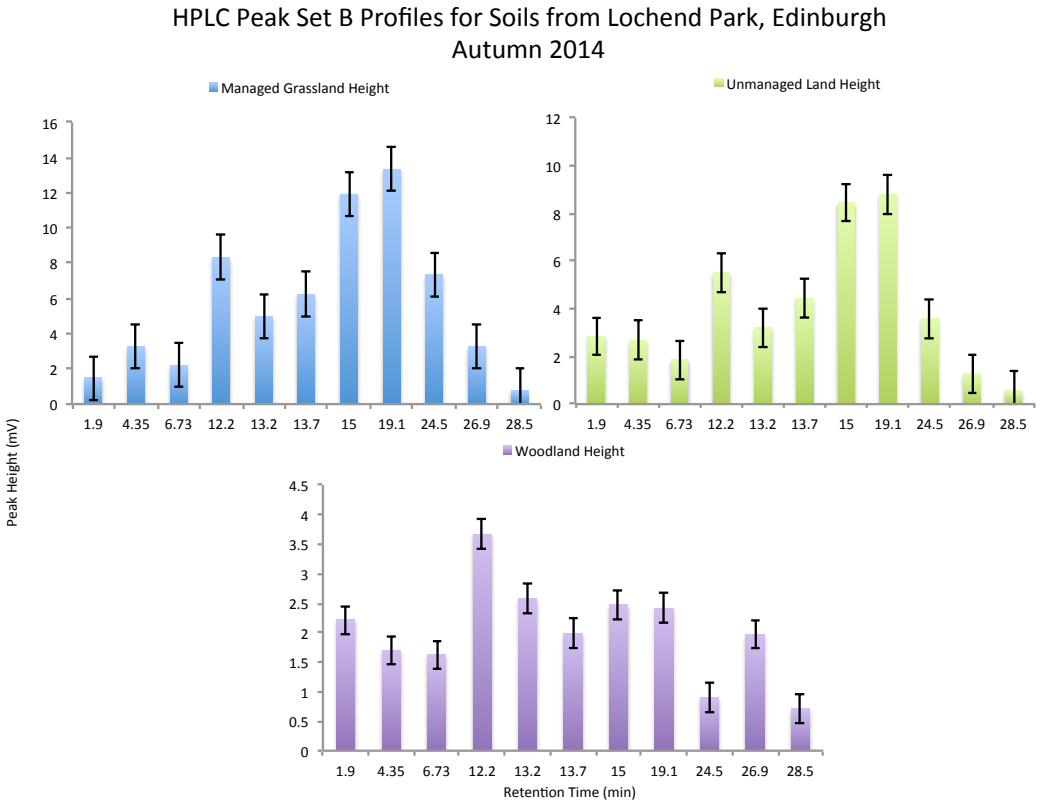
Figure 6.14 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set B- Edinburgh, Summer.



Autumn 2014

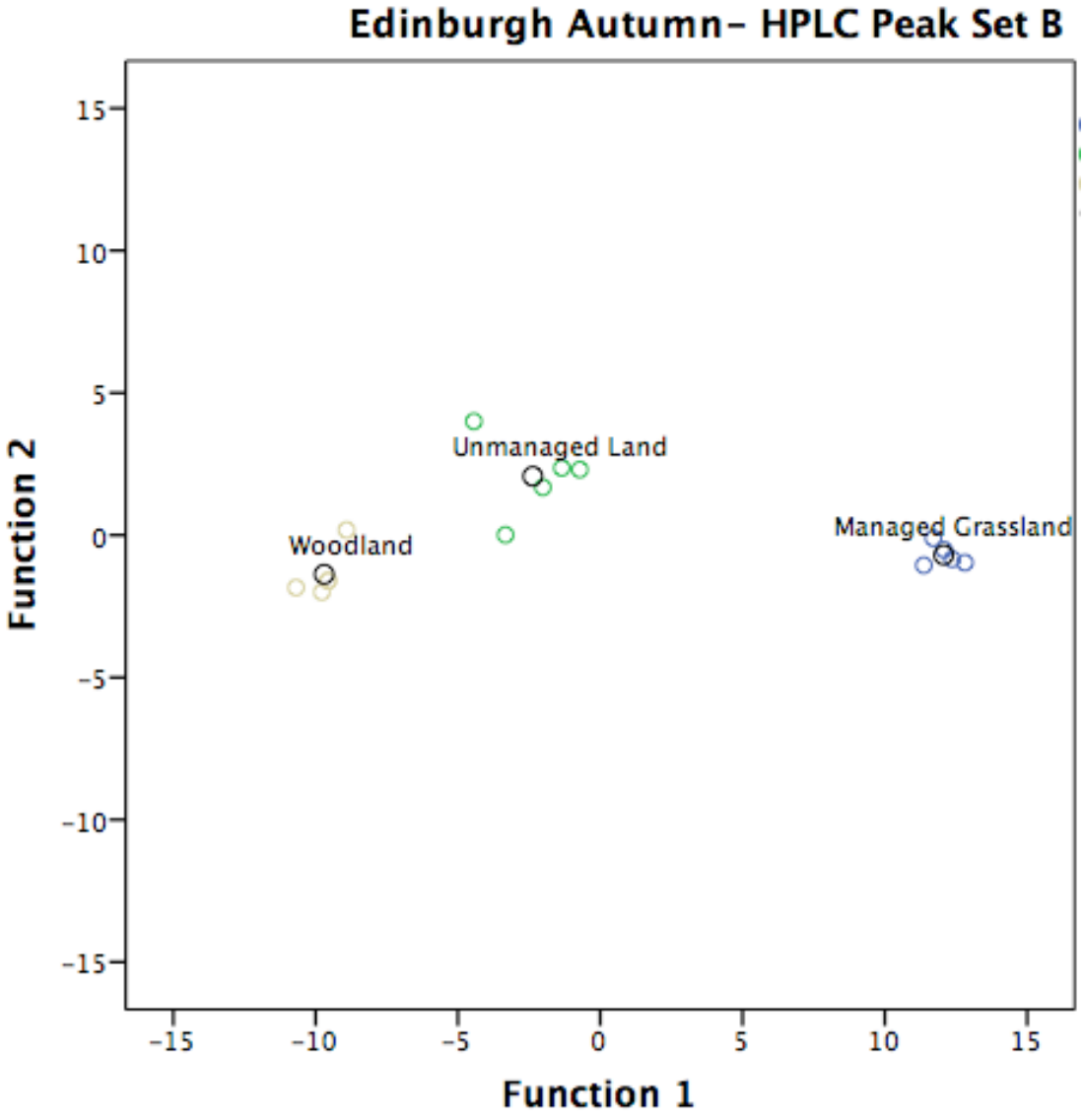
Insufficient sample aliquots could be extracted from the porous soil at the location adjacent to fresh water at this time point, therefore no chromatograms were obtained for these samples. The woodland soils were more easily characterised, visually (Figure 6.15), since the primary peak was at 12.2min and was 49% larger than the next largest peak at 15min, whereas it was 30% smaller for managed grassland and 35% smaller for unmanaged land. The ratio of the peaks at 15min to those at 1.9min allowed the managed and unmanaged land to be discriminated on the basis of visual assessment as the ratio of the two peaks was 8:1, respectively for managed grassland and only 3:1, respectively for unmanaged land.

Figure 6.15 Seasonal Changes to Peak Set B Profiles- Edinburgh, Autumn



The discrimination produced in the CDFA (Figure 6.16, Table 6.3) was 100% accurate in grouping the samples and was statistically significant at the 99% confidence level (p=0.000). The two resultant functions accounted for 97.4% and 2.6% of the variation between the three sample groups.

Figure 6.16 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set B- Edinburgh, Autumn.



6.3.2 Aberdeen

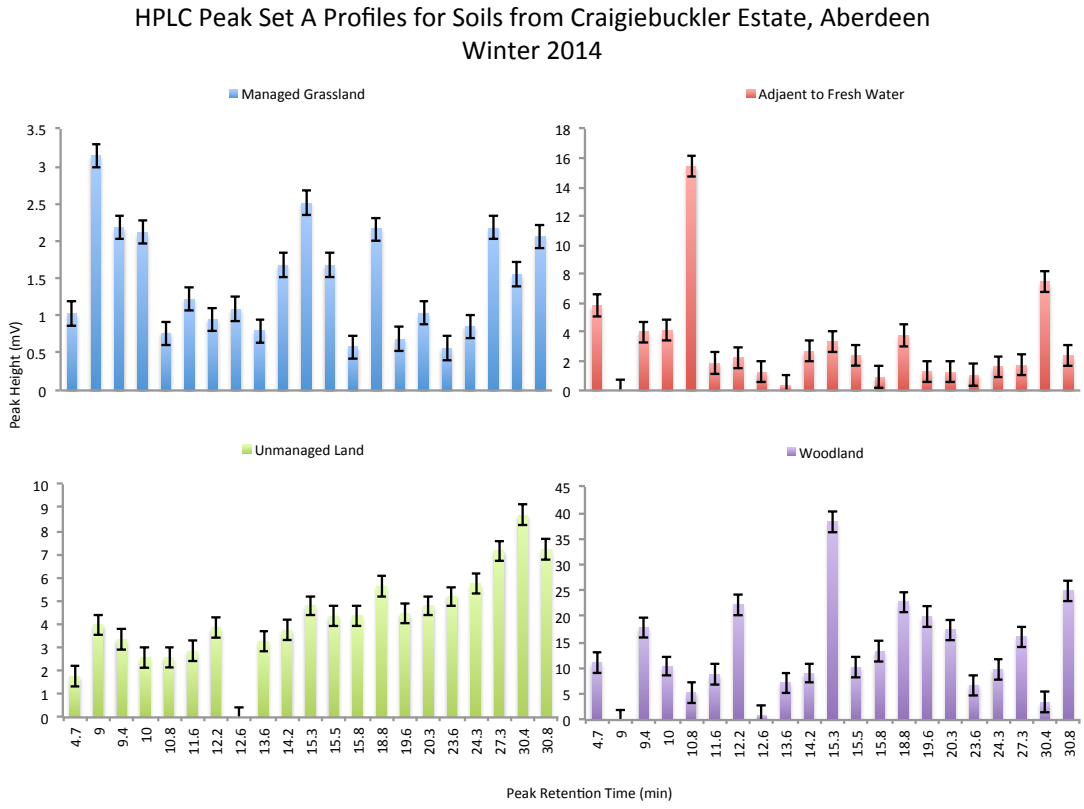
6.3.2.1 HPLC Profiles for Peak Set A

The results of the visual comparison of the Peak Set A profiles obtained at Craigiebuckler, Aberdeen and the results of the CDFA using the peaks in set A as variables, are summarised below for the four 2014 time points detailed in Table 6.1.

Winter 2014

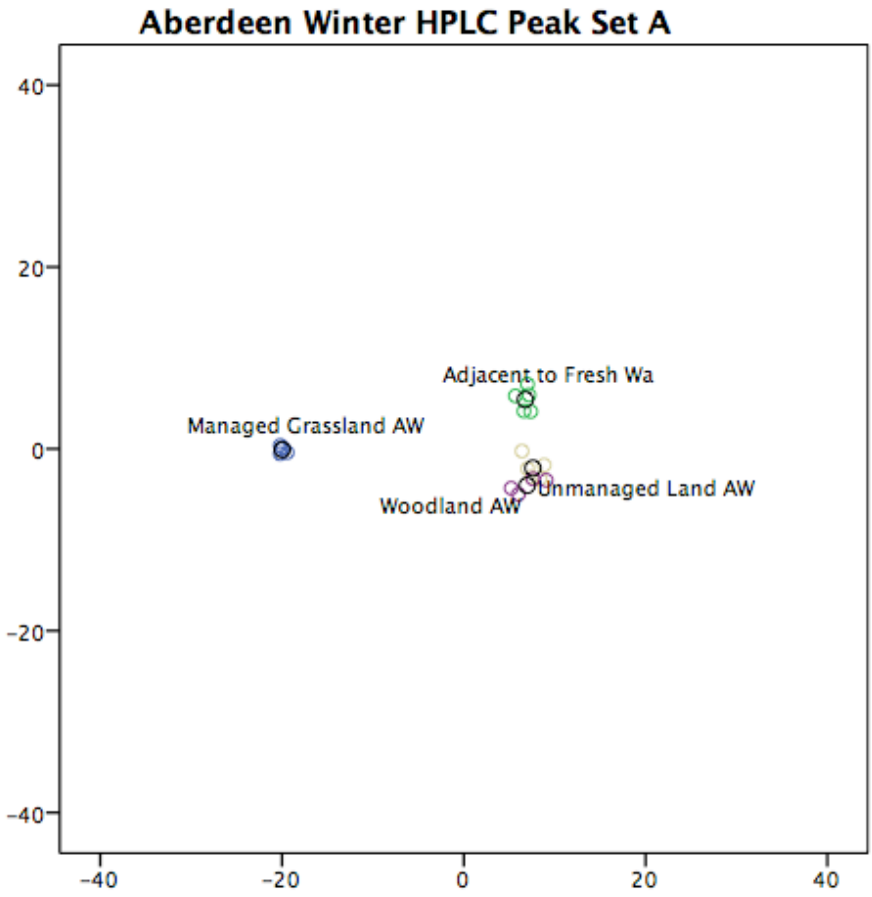
As discussed in Chapter 5, it was possible to visually discriminate the profiles for HPLC peak set A allowed each of the four locations in Craigiebuckler Estate, Aberdeen (Figure 6.17). The largest peak present in The managed grassland samples could be discriminated by having their largest peak at 9.4min, which was distinct from the other three locations. For the profiles of soils adjacent to fresh water, the largest peak was at 10.8min, which also differed from the other locations, and these profiles were also distinct from the managed grassland and unmanaged land through the absence of a peak at 9min. Woodland soil profiles were discernible from the unmanaged land and managed grassland through the absence of the peak at 9min and the large relative height of the peak at 30.8min. The highest peak in the profile of the samples from unmanaged land was at 30.4min, and these samples could also be discriminated by presence of the peak at 9 min and the absence of a peak at 12.6min.

Figure 6.17 Seasonal Changes to Peak Set A Profiles- Aberdeen, Winter



The discrimination produced in the CDFA (Figure 6.18, Table 6.2) was 94.7% accurate in grouping the samples and was statistically significant at the 99% confidence level (p=0.001). The resultant functions accounted for 90.6% ,7.8%, and 1.5% of the variation between the three sample groups. One sample from the unmanaged location was misclassified as a woodland soil by these functions.

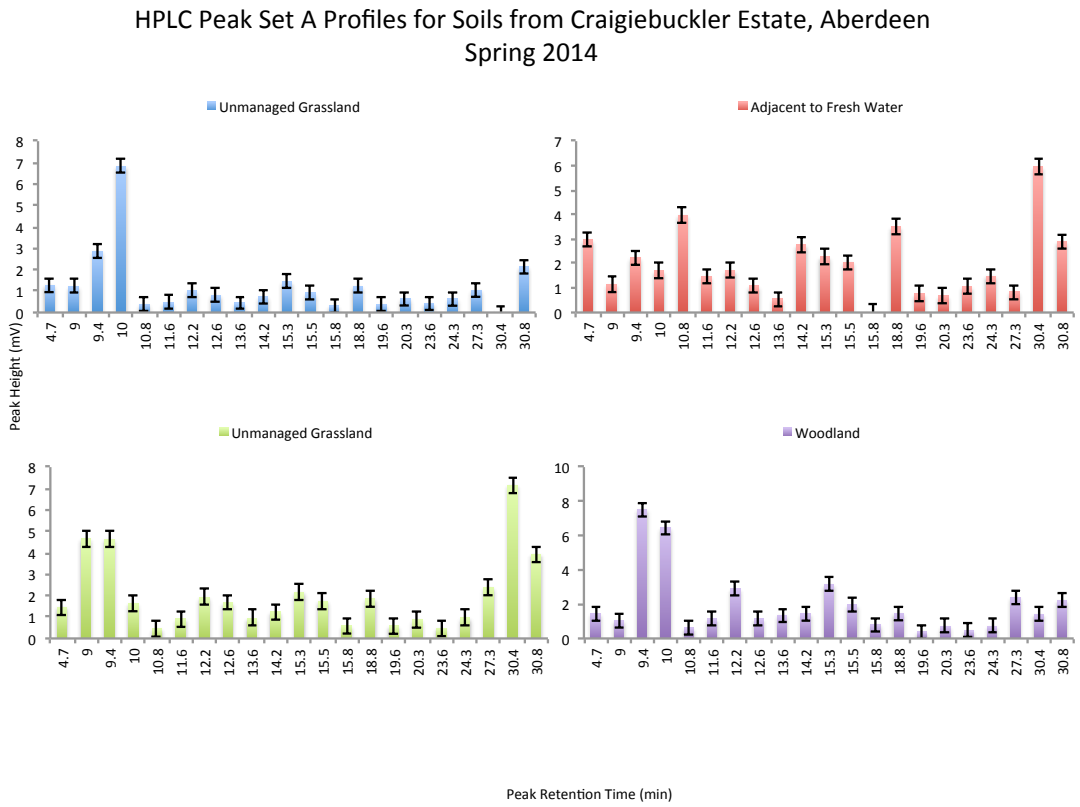
Figure 6.18 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set A- Aberdeen, Winter.



Spring 2014

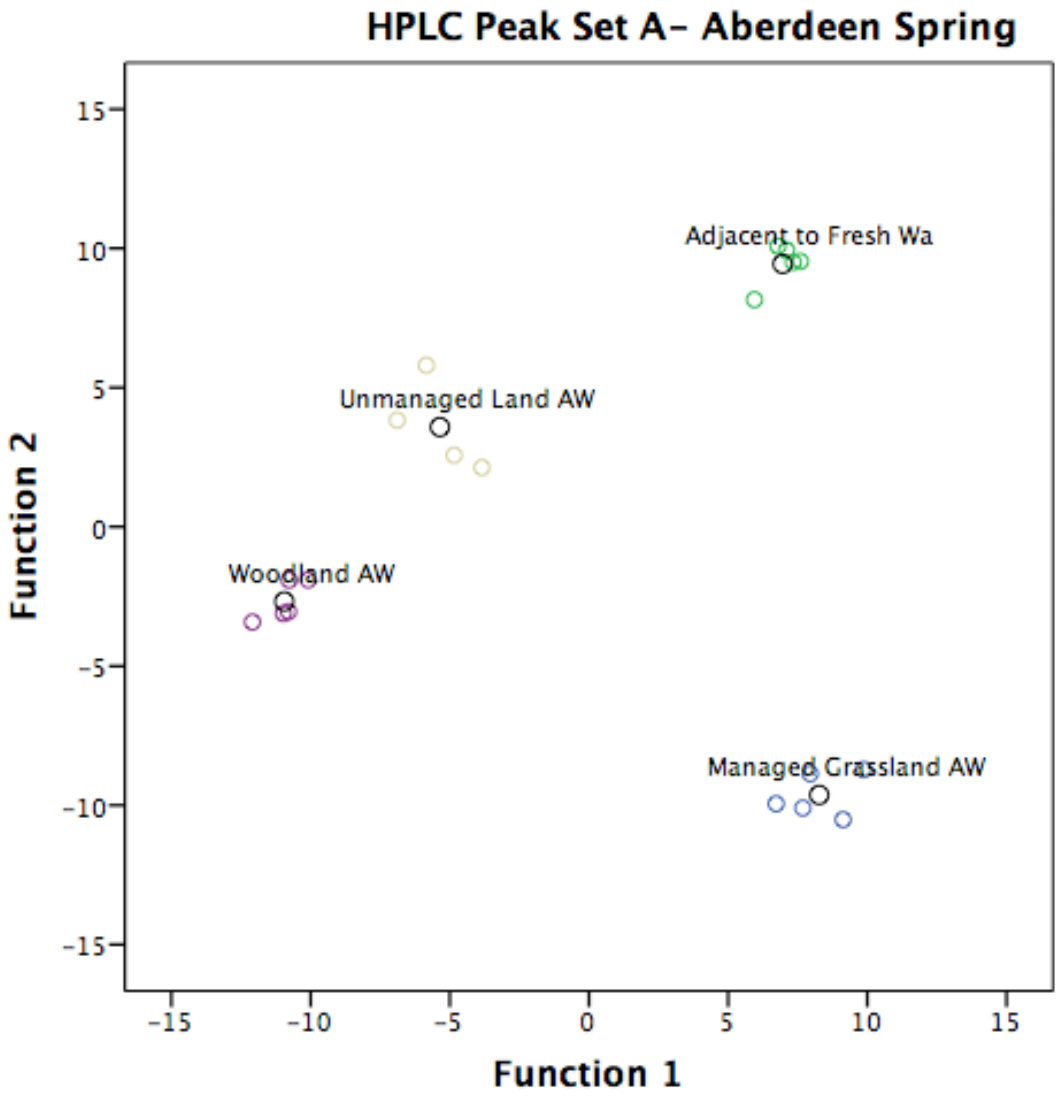
The profiles for each of the locations at the Craigiebuckler site were visually distinct (Figure 6.19), and each location had large, major components that were different from every other location. The location adjacent to fresh water had a major peak at 30.4min that was 5.95mV, and was 50% larger than the next largest peak at 10.8min. Unmanaged land also produced profiles where the peak at 30.4min was the largest peak present, at 7.16mV, however in this case the next largest peaks were those at 9min and 9.4min, which were both 35% smaller than the major peak, at 4.65mV and 4.62mV, respectively. The major peak for managed grassland was at 10min, and at 6.83mV was more than twice the size of the second largest peak, at 9.4min, which was 2.86mV. Woodland soils produced profiles for peak set B that where the major peaks were at 9.4min and 10min which, at 7.5mV and 6.5mV, were significantly larger than the remaining peaks, which ranged from 0.39mV to 3.13mV.

Figure 6.19 Seasonal Changes to Peak Set A Profiles- Aberdeen, Spring



The statistical analysis (Figure 6.20, Table 6.2) produced three canonical discriminant functions which explained 45.0%, 34.6% and 20.4% of the variance in the samples, respectively and the samples were grouped with 100% accuracy using these three functions. Furthermore, the discrimination using these profiles was statistically significant at the 99% confidence interval (p=0.000)

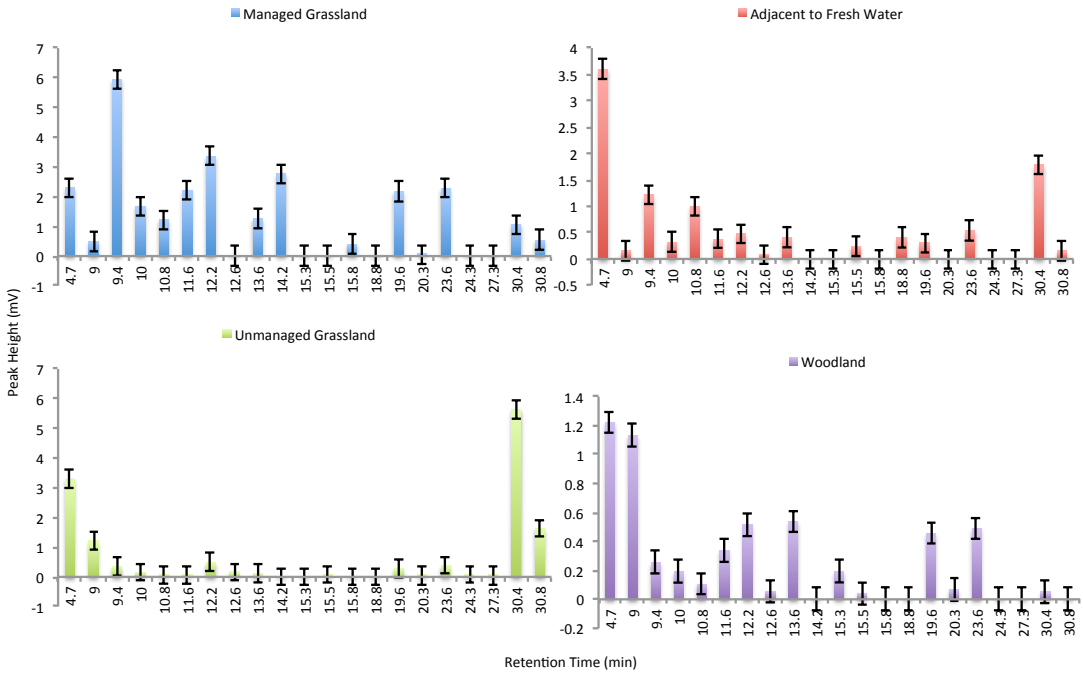
Figure 6.20 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set A- Aberdeen, Spring.



Summer 2014

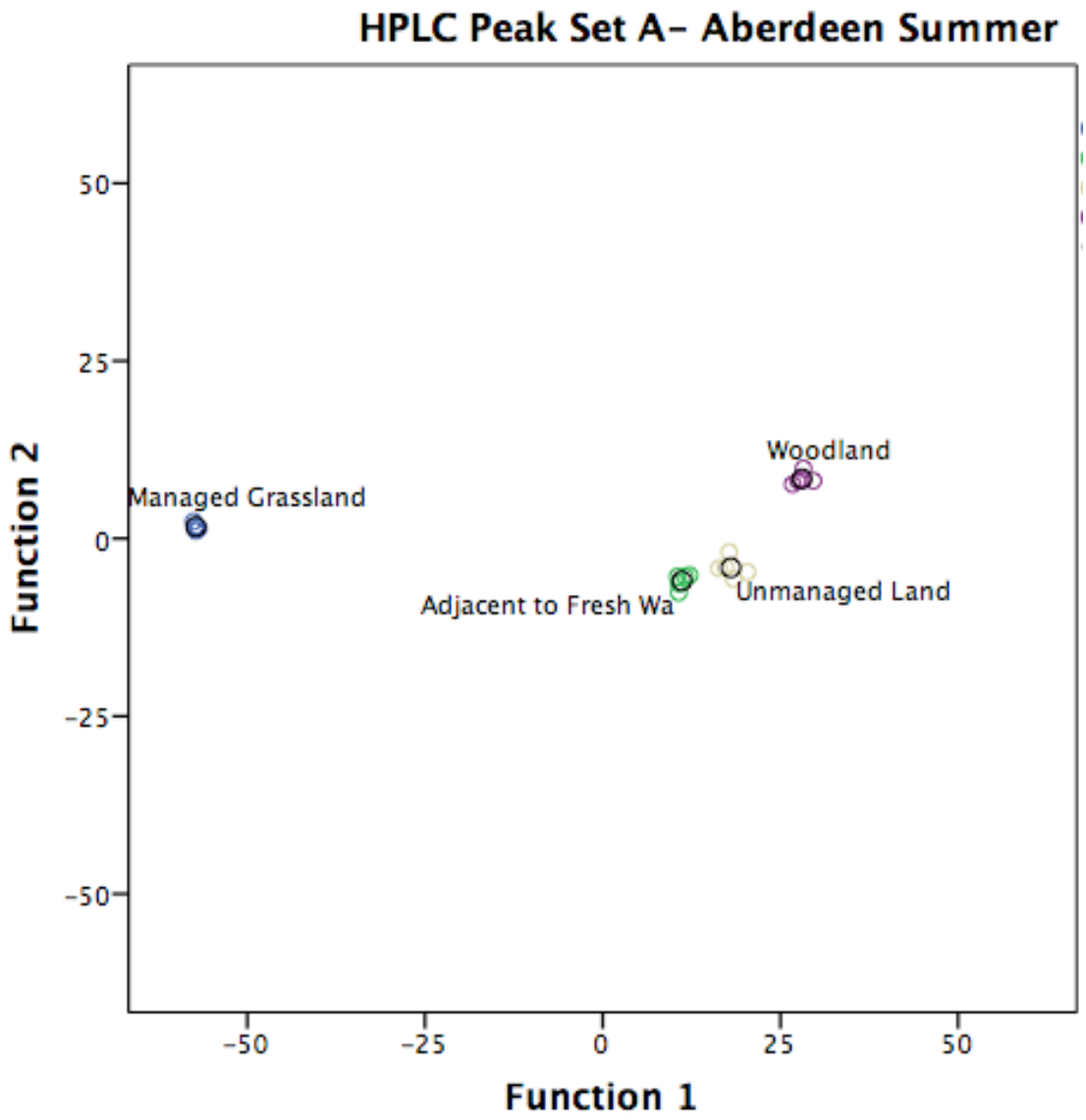
In Summer, the peak set A profiles for Aberdeen were could be discriminated visually (Figure 6.21), with the two largest peaks in each profile being a simple method of telling the locations apart from one another. For woodland, the two largest peaks were at 4.7min and 9min, at 1.2mV and 1.1mV, respectively, while for unmanaged land the two largest peaks were at 30.4min and 4.7min with peak heights of 5.6mV and 3.3mV. For the soils adjacent to fresh water the two largest peaks were also 4.7min and 30.4min, however the size order was reversed and the peaks were 3.6mV and 1.8mV, respectively. The two largest peaks in the profiles obtained for managed grassland were at 9.4min and 12.2min with heights of 5.9mV and 3.4mV, respectively.

Figure 6.21 Seasonal Changes to Peak Set A Profiles- Aberdeen, Summer



The statistical analysis using peak set A produced three canonical discriminant functions (Figure 6.22, Table 6.2) which explained 96.8%, 2.7% and 0.5% of the variance in the samples, respectively, and gave 100% accuracy in grouping the samples using these functions, and this discrimination was significant at the 99% confidence interval (p= 0.000) when all three functions were used together. The separation achieved using the first two canonical functions is displayed in the scatter plot in Figure 6.22.

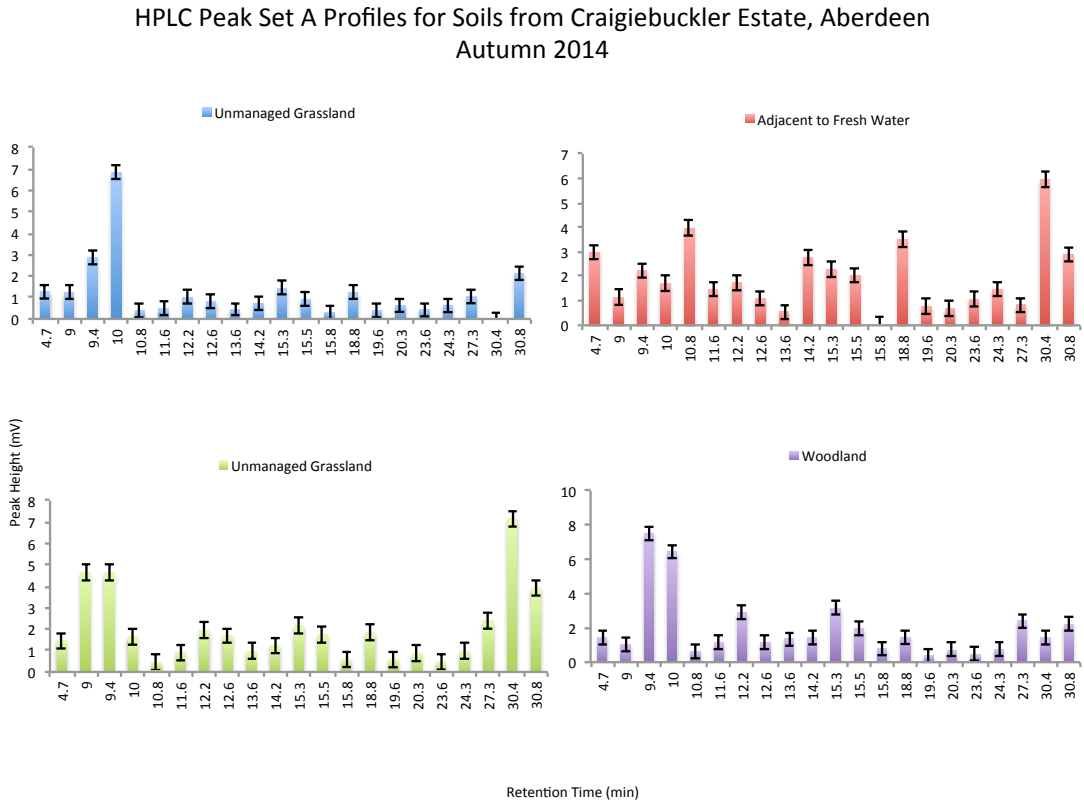
Figure 6.22 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set A- Aberdeen, Summer.



Autumn 2014

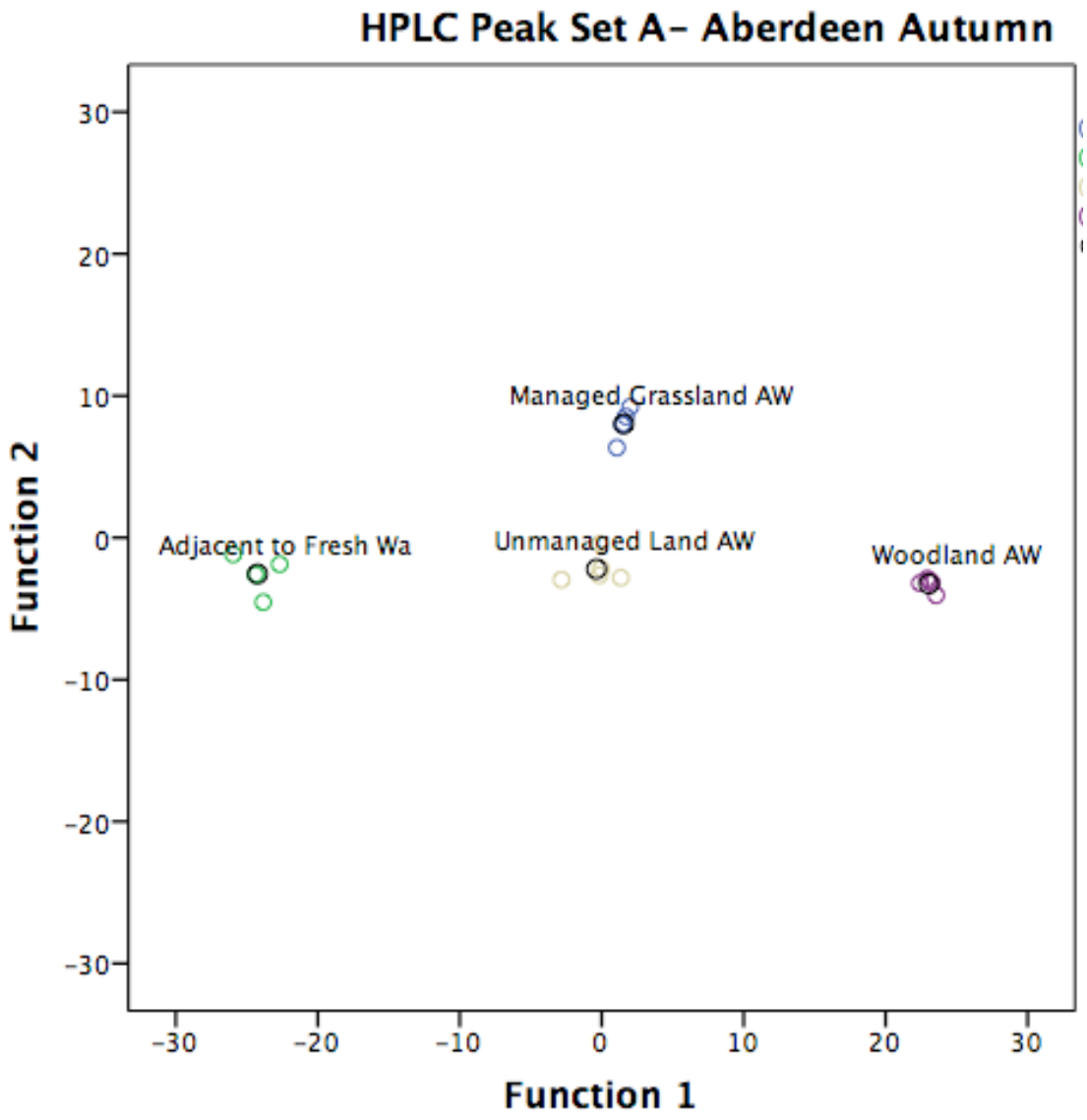
The profiles for the Craigiebuckler site in Autumn displayed mostly relatively small peaks, and the comparison of the largest two peaks in the profiles was sufficient to classify the samples in this data set (Figure 6.23). For managed grassland the largest two peaks were at 10min with a height of 6.8mV, and 9.4min with a height of 2.9mV. For the soils adjacent to fresh water, the two tallest peaks were at 30.4min, where the peak was 5.9mV, and at 10.8min where the peak height was 4.0mV. For the soils from the unmanaged locations 30.4min where the peak was 7.2mV, and 9 min and 9.4min where the peak heights were both 4.6mV. The woodland soils had two large peaks at 9.4min and 10min, which were at 7.5mV and 6.5mV in height

Figure 6.23 Seasonal Changes to Peak Set A Profiles- Aberdeen, Autumn



The CDFA (Figure 6.24, Table 6.2) provided 100% accuracy using these profiles, and this discrimination was statistically significant at the 99% confidence level (p=0.000) . The three functions accounted for 90.5%, 7.0%, and 2.5% of the variance in the sample groupings.

Figure 6.24 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set A- Aberdeen, Autumn.



6.3.2.2 HPLC Profiles for Peak Set B

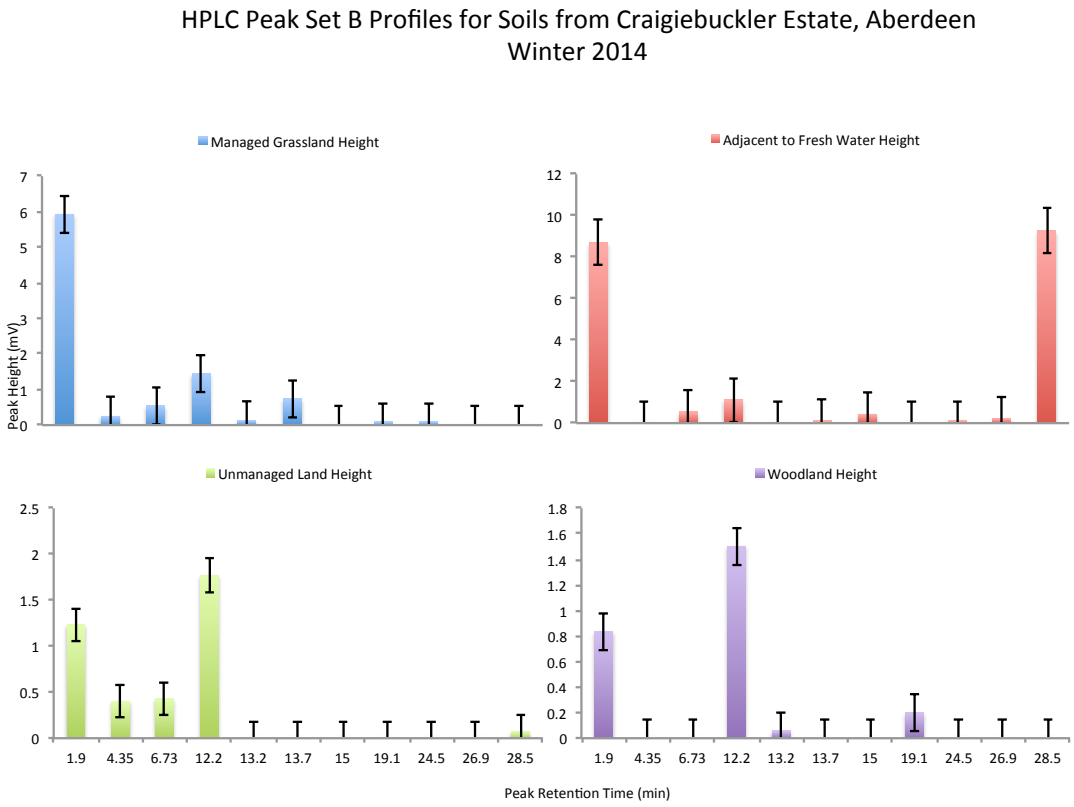
This section outlines the results of the visual comparison of the Peak Set B profiles obtained at Craigiebuckler, Aberdeen and the results of the CDFA using the peaks in set B as variables for the four 2014 time points detailed in Table 6.1.

Winter 2014

As discussed in Chapter 5, the samples from the four locations within the Craigiebuckler Estate could clearly be distinguished through visual comparison of peak set B profiles in Winter (Figure 6.25). The retention time of and ratio between the two largest peaks at each location were useful features for discriminating the samples from soils adjacent to fresh water, which had two large, similarly sized peaks at 1.9 and 28.5min, from the managed grassland samples,

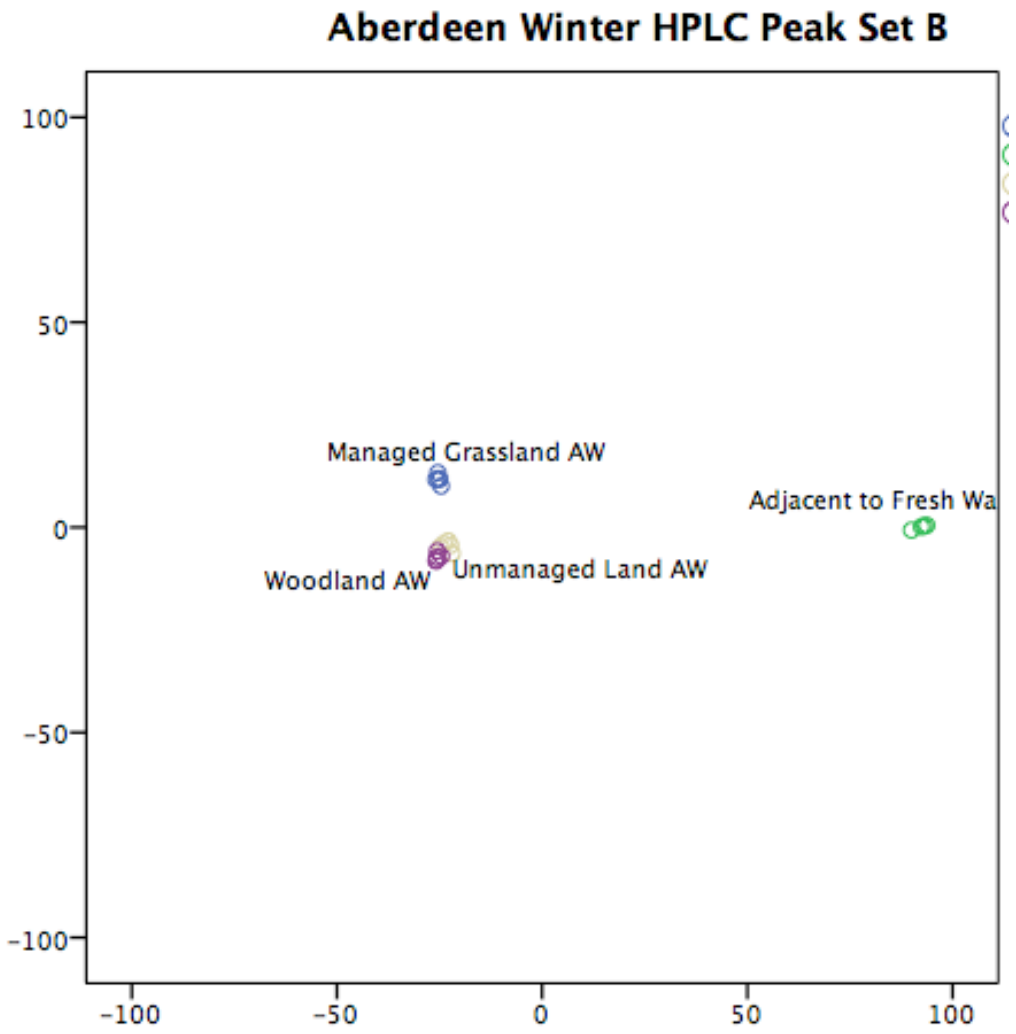
for which the largest peak was at 1.9min and next largest at 12.2min. The ratio of the largest peak at 12.2min to the next largest peak at 1.9min was greater for woodland soils at 8:1, than for the unmanaged location at 3:1, and woodland soils were also missing the peaks at 4.35 and 6.73min that were present at the unmanaged location.

Figure 6.25 Seasonal Changes to Peak Set B Profiles- Aberdeen, Winter



In Winter, 100.0% accuracy was achieved when grouping the Aberdeen samples using the three functions produced in the CDFA for peak set B (Figure 6.26, Table 6.3), which explained 97.4%, 2.4% and 0.2% of the variation in the samples, and this discrimination was statistically significant at the 99% confidence interval.

Figure 6.26 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set B- Aberdeen, Winter.

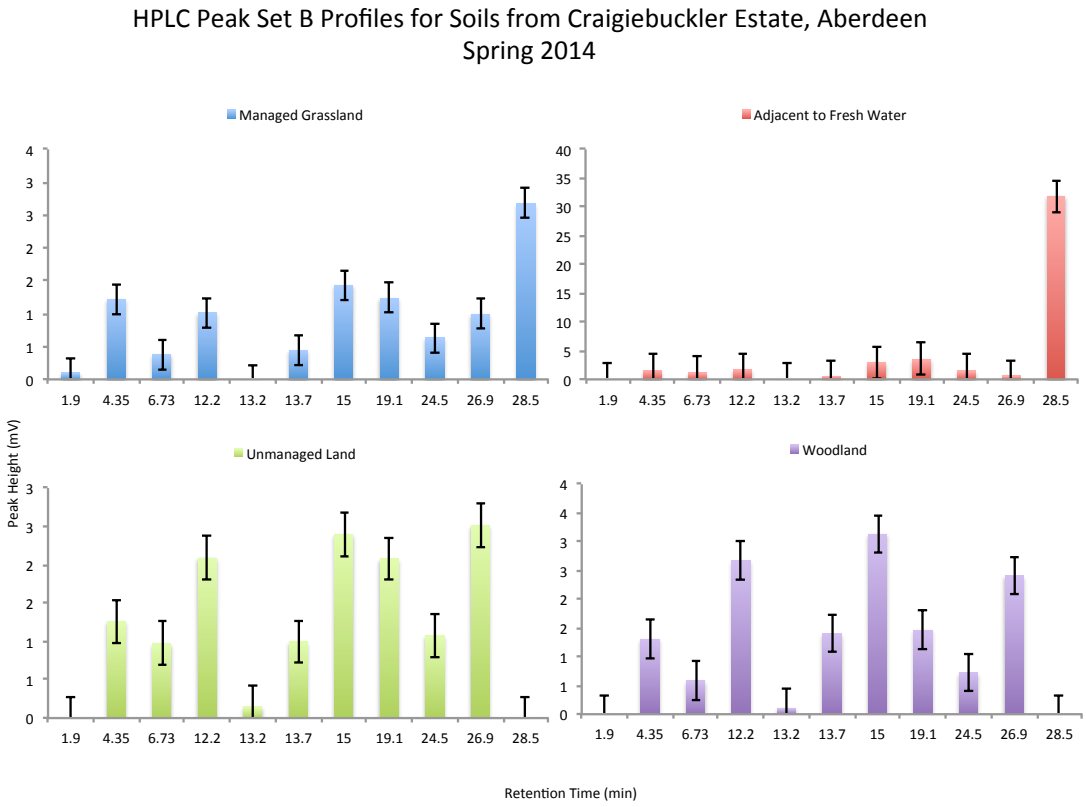


Spring 2014

There were few large peaks in the peak set B profiles for the location adjacent to fresh water at the spring time point at the Aberdeen site (Figure 6.27), and all but one of the peaks present were in the range of 0.5mV to 3.5mV, therefore these soils were easily discriminated from the other locations by the presence of single large peak of 31.9mV at 28.5 min. The managed grassland soil profiles also had their main peak at 28.5min, with an average height of 2.9mV, however these soils were easy to visually separate from the soils adjacent to fresh water due to the relative sizes of the other peaks, which were in the range of 0.1-1.4mV, so the ratio of the peak at 28.5min to any other peak was much lower for managed soils, for instance the ratio of the peak at 28.5min to the peak at 15min was 1.8:1, whereas the same ratio was 10.6:1. The peak at 28.5min was absent from the profiles of the soils from both unmanaged

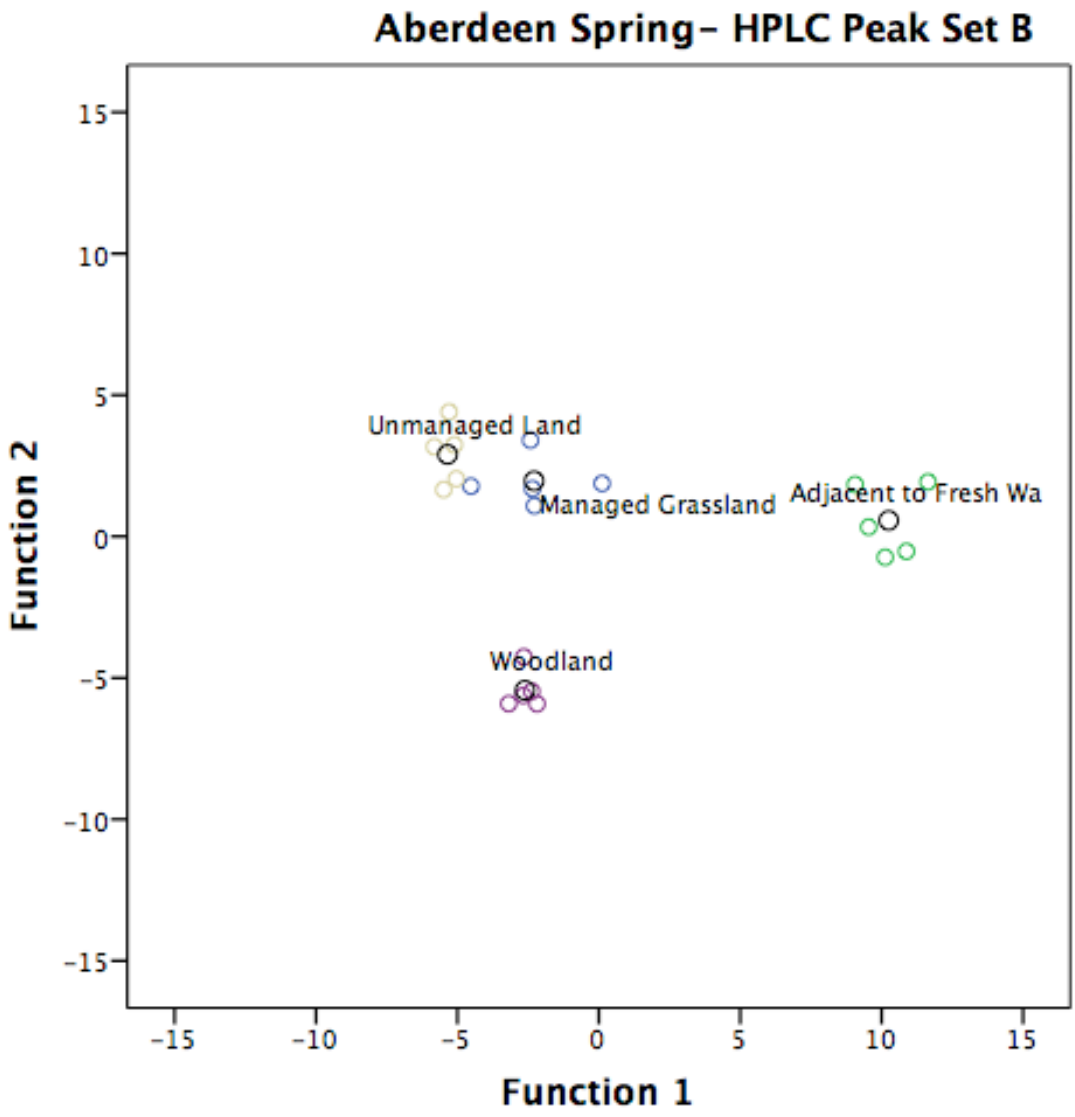
land and woodland, and these two locations could be discriminated from one another through comparison of the relative sizes of the peaks at 15min and 19.1min, since the peak at 15min was 15% larger than the peak at 19.1min for the unmanaged land but 113% larger for the woodland soils.

Figure 6.27 Seasonal Changes to Peak Set B Profiles- Aberdeen, Spring



Using the data for peak set B, the Spring samples from Aberdeen were again grouped with 100% accuracy by CDFA (Figure 6.28, Table 6.3), and the discrimination was significant at the 99% confidence interval (p=0.000) The three functions produced in the CDFA were responsible for 74.9%, 21.7%, and 3.5% of the variability in the groupings

Figure 6.28 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set B- Aberdeen, Spring

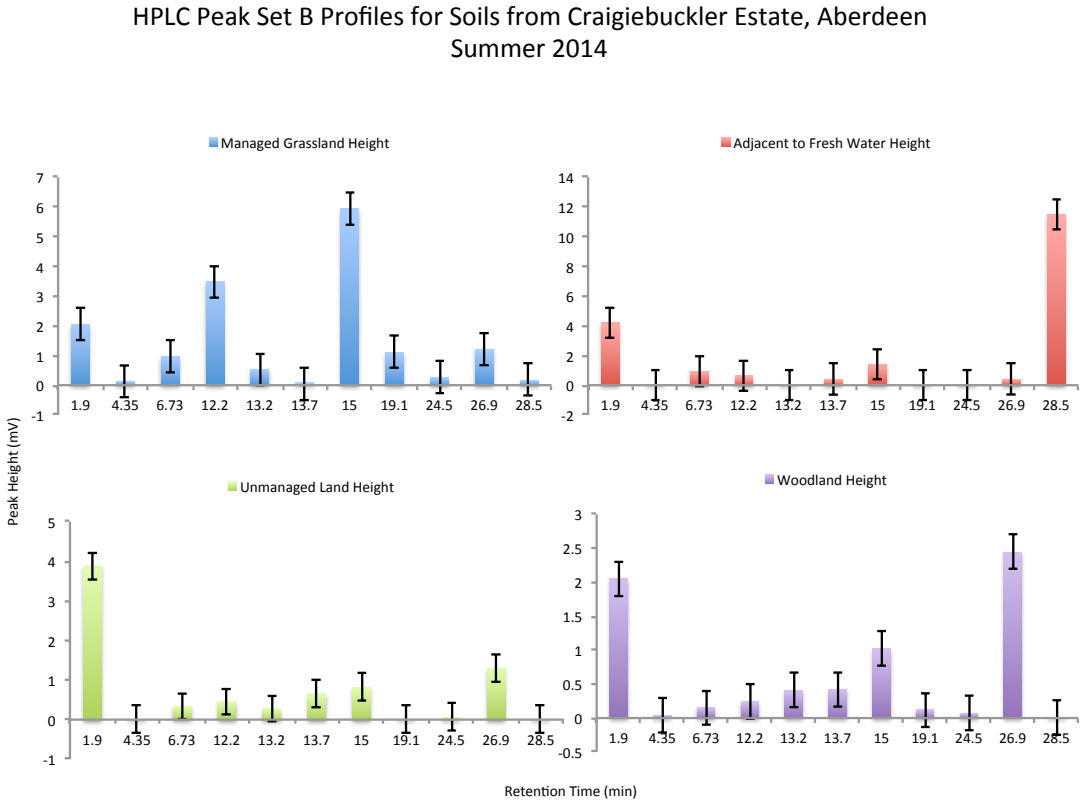


Summer 2014

The peak set B profiles for the Craigiebuckler site in Summer displayed 2-3 major peaks at each location which allowed the different locations to be distinguished (Figure 6.29). For the managed grassland, the profiles could be identified by three prominent peaks at 1.9min, 12.2min and 15min, which were 2.0mV, 3.5mV and 5.9mV in height, while all other peaks were in the range 0.2mV to 1.2mV. The soils from locations adjacent to fresh water were characterised by two large peaks at 1.9min and 28.5min, at 4.2mV and 11.5mV in height respectively, while all other quantifiable peaks present were in the range 0.4mV to 1.4mV. The profiles from woodland soils and those from unmanaged land were both characterised by having their two largest peaks at 1.9min and 26.9min, however the size order was reversed

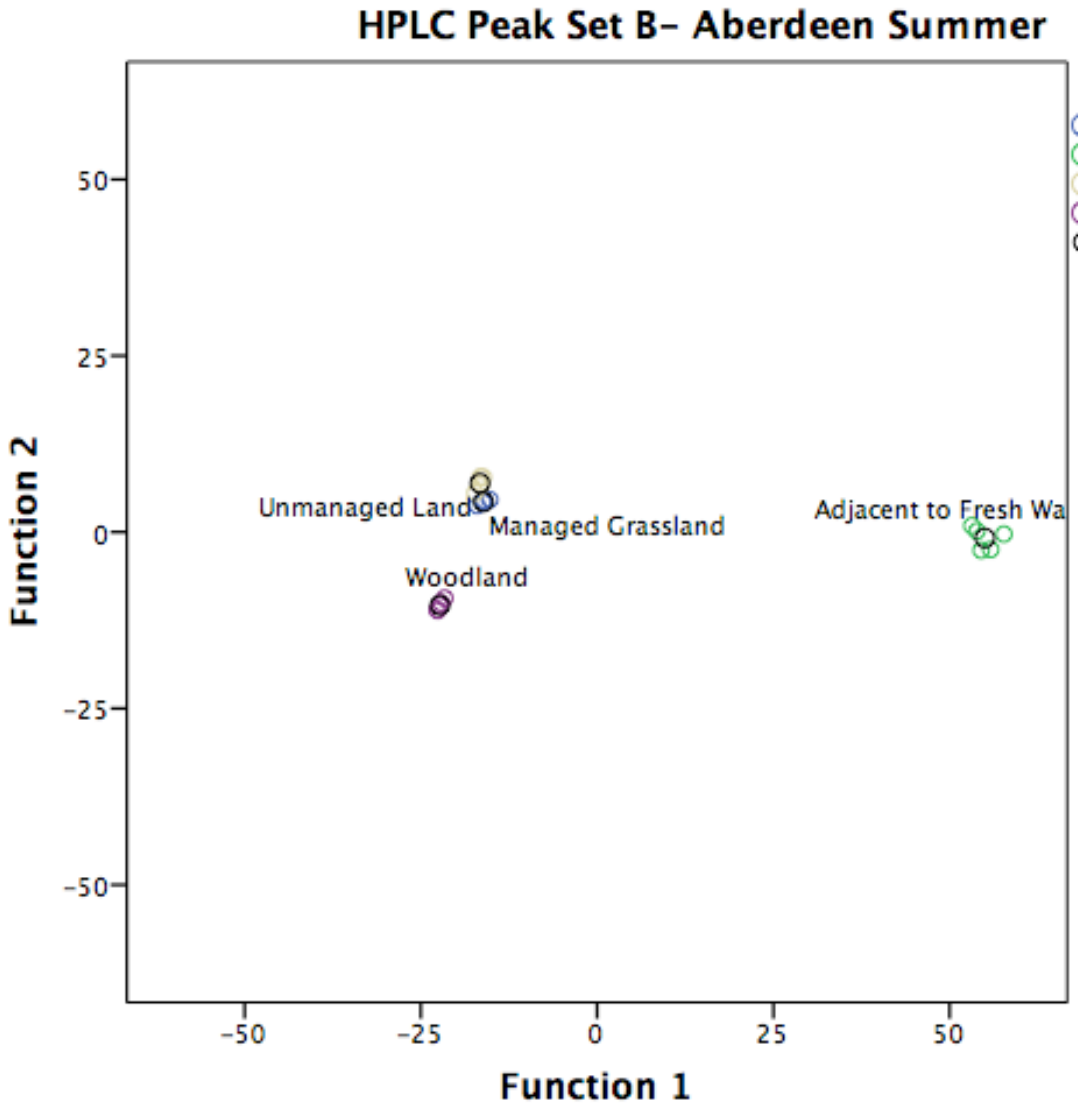
between the two locations; the peak at 1.9min was 200% larger than the peak at 26.9min for the unmanaged land but was 19% smaller for woodland soils.

Figure 6.29 Seasonal Changes to Peak Set B Profiles- Aberdeen, Summer



The CDFA (Figure 6.30, Table 6.3) on the data for peak set B also produced functions that afforded 100% accuracy in grouping the samples from Aberdeen in Summer, which accounted for 95.4%, 4.1%, and 0.5% of the difference between the sample groups. The discrimination produced for this data set was statistically significant at the 99% confidence interval (p=0.000)

Figure 6.30 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set B- Aberdeen, Summer

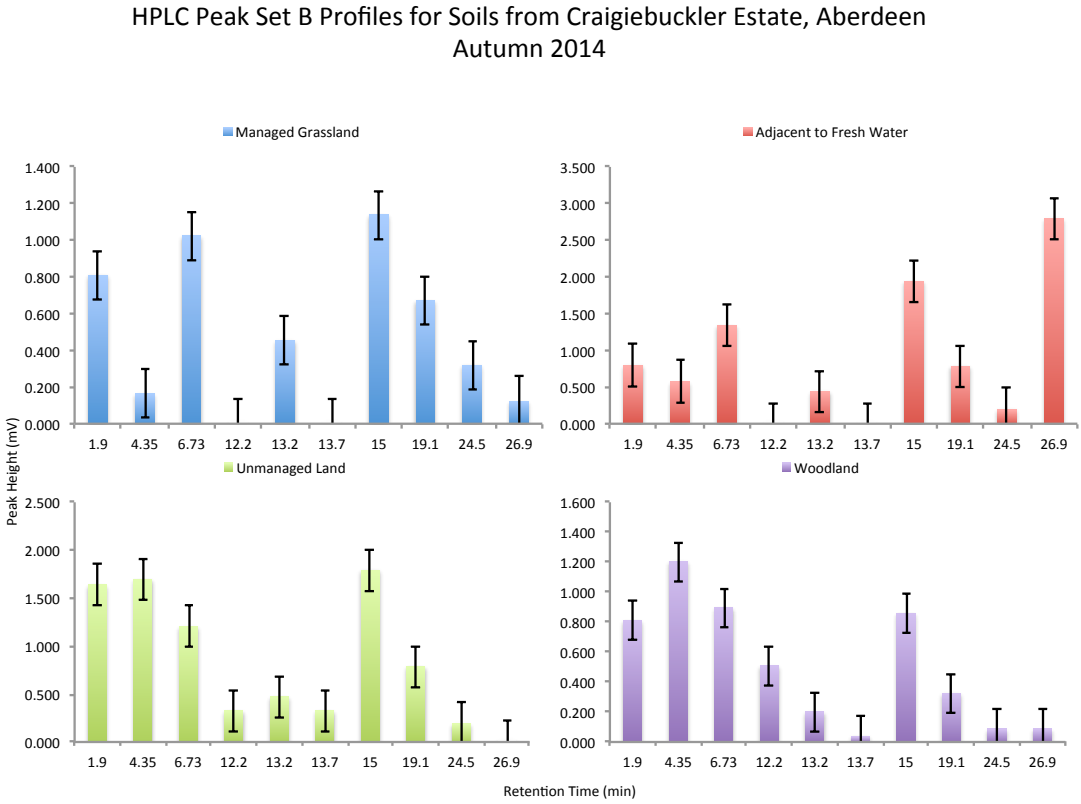


Autumn 2014

At the Autumn time point (Figure 6.31), the peak set B profiles for the Craigiebuckler site could be separated into two groups, since the peaks at 12.2min and 13.7min were absent from the profiles of the soils from both managed grassland and locations adjacent to fresh water, but present in the woodland soils and those from unmanaged land. The locations adjacent to fresh water and the managed grassland profiles could be separated by the relative sizes of the peaks at 15min, which at 1.1mV was the largest peak for the managed grassland soils, and the peak at 26.9min, which at 2.8mV was the largest peak for the locations adjacent to fresh water. The earlier eluting peak was 89% smaller than the later eluting peak for the managed grassland while it was 44% larger for the soils adjacent to fresh water. In addition, the unmanaged land and woodland soils could be distinguished since the peaks at 4.35min and

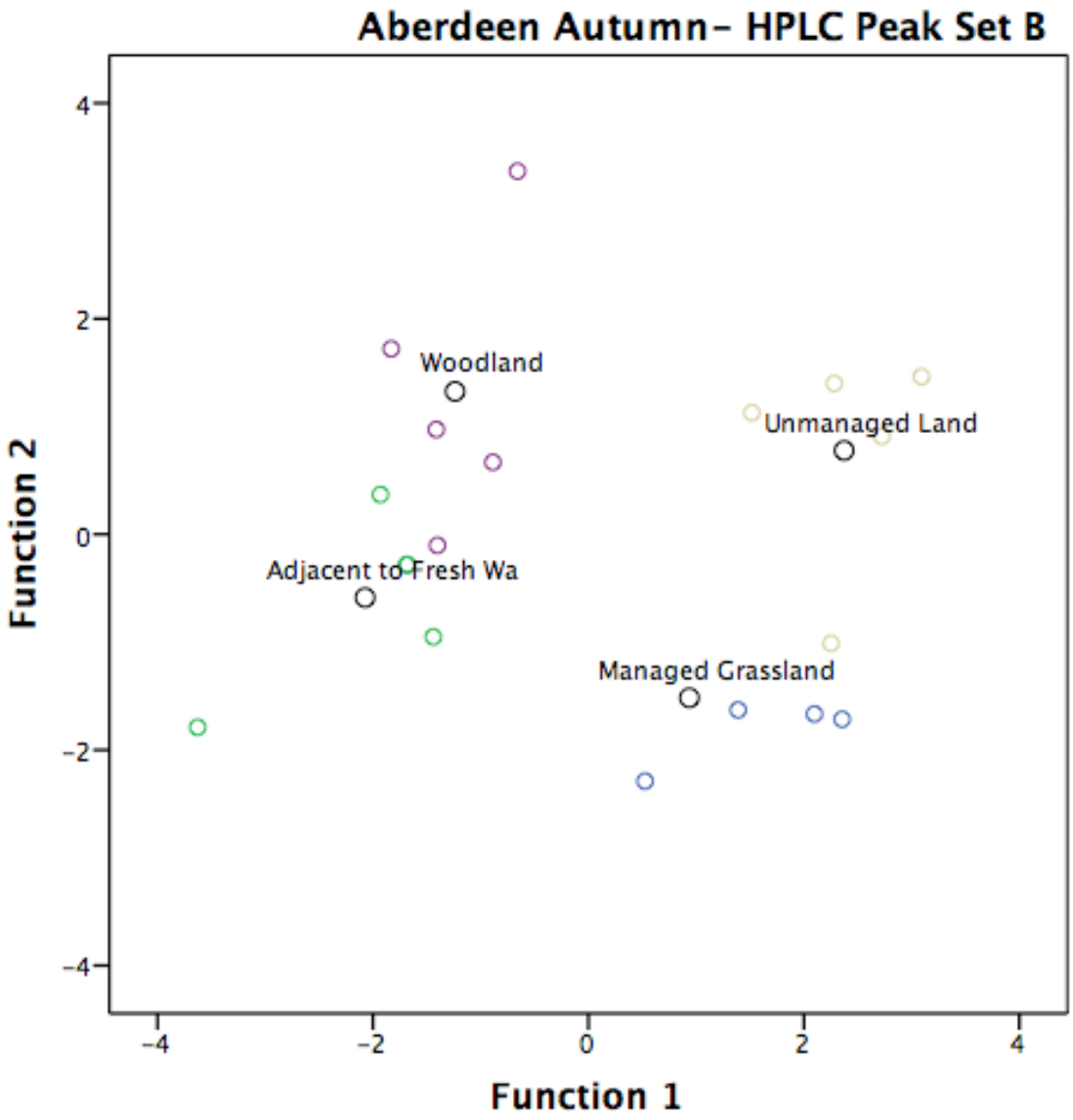
15min were approximately equal for the unmanaged land, at 1.70mV and 1.79mV respectively, while for woodland the peak at 4.35min was 39% larger than the peak at 15min.

Figure 6.31 Seasonal Changes to Peak Set B Profiles- Aberdeen, Autumn



The accuracy of sample groupings for the peak set B data was 95% for Aberdeen in Autumn, as one of the managed grassland samples was misclassified as having originated from the location adjacent to fresh water using the first three canonical functions, which produced 60.7%, 24.5% and 14.7% of the variation in the data. This discrimination, however, was not statistically significant (p=0.145). The CDFA scatter plots for this data set is presented in figure 6.32.

Figure 6.32 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set B- Aberdeen, Autumn



6.3.3 London

6.3.3.1 HPLC Profiles for Peak Set A

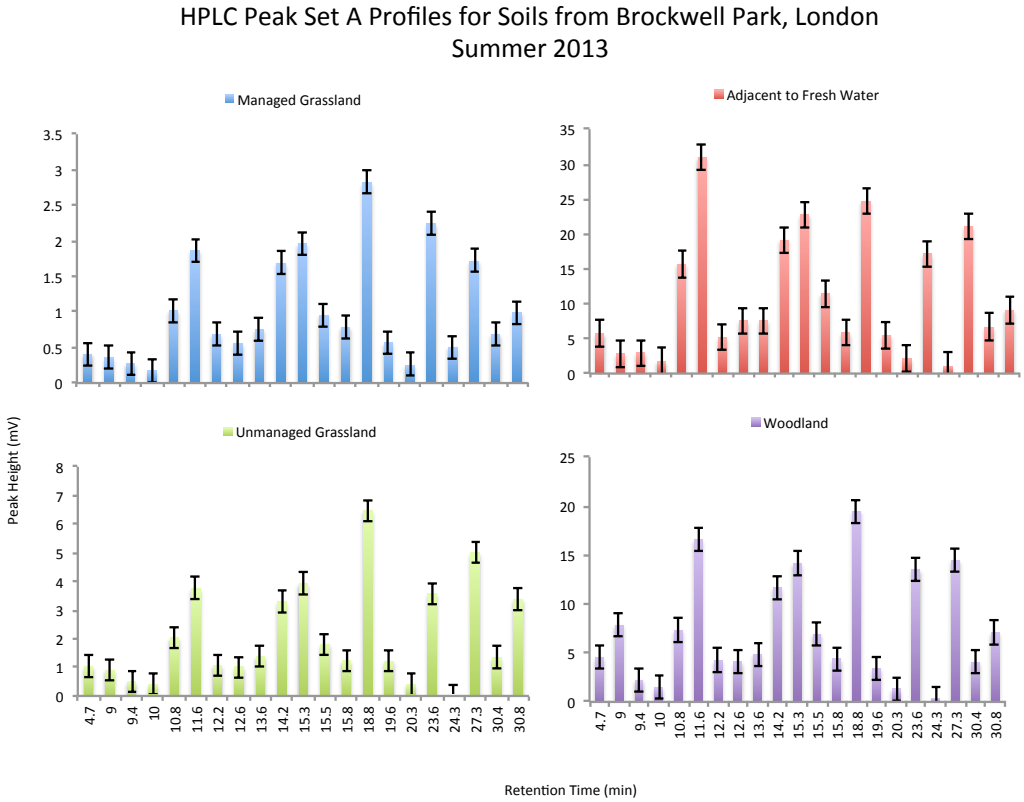
The results of the visual comparison of the Peak Set A profiles obtained at Brockwell Park, London are presented below, alongside the results of the CDFA using the peaks in set A as variables for the six time points detailed in Table 6.1

Summer 2013

The profiles for the soils adjacent to fresh water were distinct in having the peak at 11.6min as the most prominent peak, at 30.9mV, compared to the next tallest peak at 18.8min, which was 24.7mV in height, while the peak at 18.8min was the largest peak for all other locations (Figure

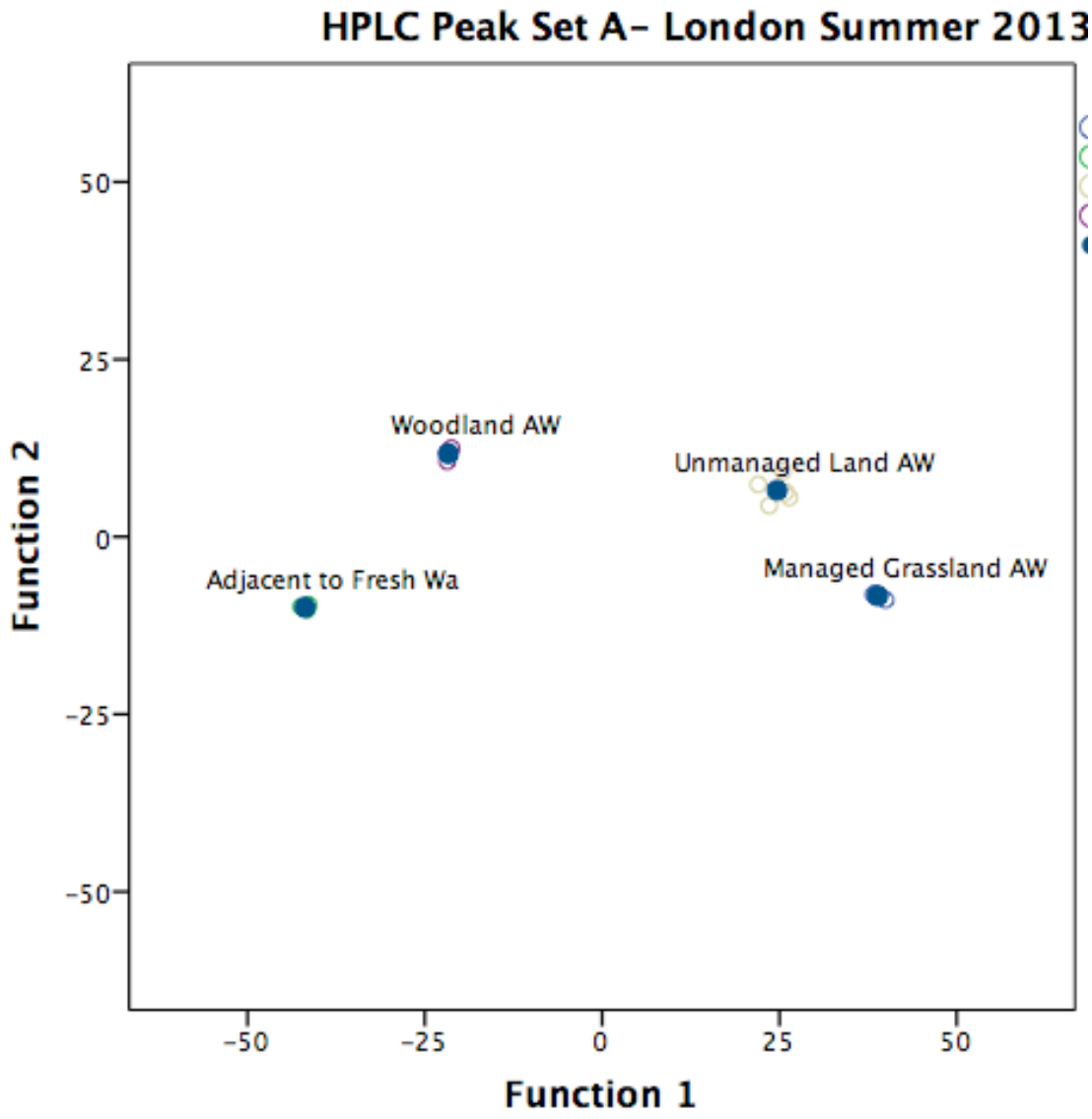
6.33). Managed grassland was distinct from both woodland soils and those from unmanaged land in that the peak at 23.6min was 30% larger, at 2.24mV, than the peak at 27.3min, which was 1.72mV, while for the woodland soils the earlier peak was 28% smaller, at 3.57mV than the later peak, which was 5.00mV in height, and for the woodland soils the earlier peak was 7% smaller than the later peak, with peak heights of 13.48mV and 14.49mV, respectively. The size difference between the peaks at 11.6min and 18.8min provided an additional way of separating the profiles of the soils from unmanaged land and woodland, since there was an increase of 18% for the former location compared to an increase of 72% for the latter.

Figure 6.33 Seasonal Changes to Peak Set A Profiles- London, Summer 2013



The CDFA (Figure 6.34, Table 6.2) for this data set grouped the samples with 100% accuracy using three canonical functions, which accounted for 91.7%, 7.4%, and 1.0% of the variation in the samples, and this discrimination was statistically significant at the 99% confidence interval (p=0.000)

Figure 6.34 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set A- London, Summer 2013

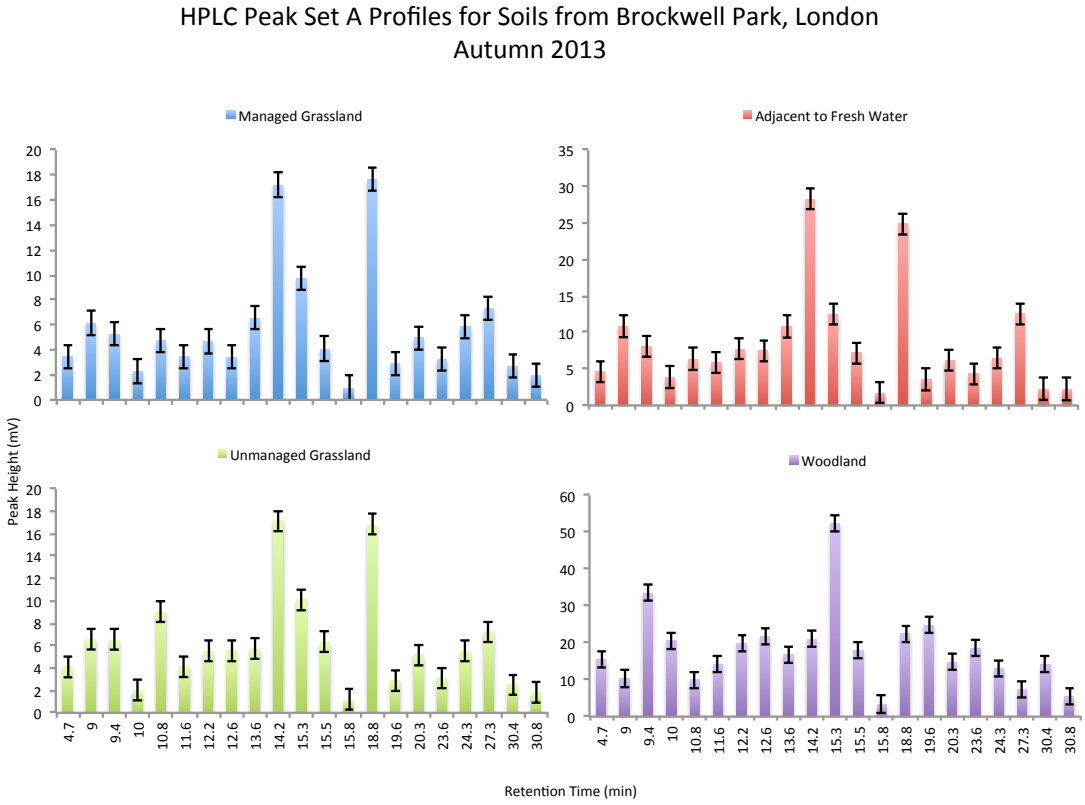


Autumn 2013

The peak set A profiles (Figure 6.35) of the woodland soils were the most obviously different from the other locations at the Autumn 2013 time point at the London site, since the peak at 15.3 was the most prominent peak, which was 56% larger in height, at 52.4mV, than the next largest peak at 9.4min, which was 33.5mV. In contrast the two most prominent peaks were at 14.2min and 18.8min at the other three locations, where the difference in size between the peak at 14.2 min and the peak at 18.8min was much smaller, ranging from -12% to 2%. The ratio of the peak height 10.8min compared to the peak height at 10min was indicative of the profiles at the unmanaged location, at 4.5:1, while the peaks were more close in size for the profiles of the soils from managed grassland and the location adjacent to fresh water, where

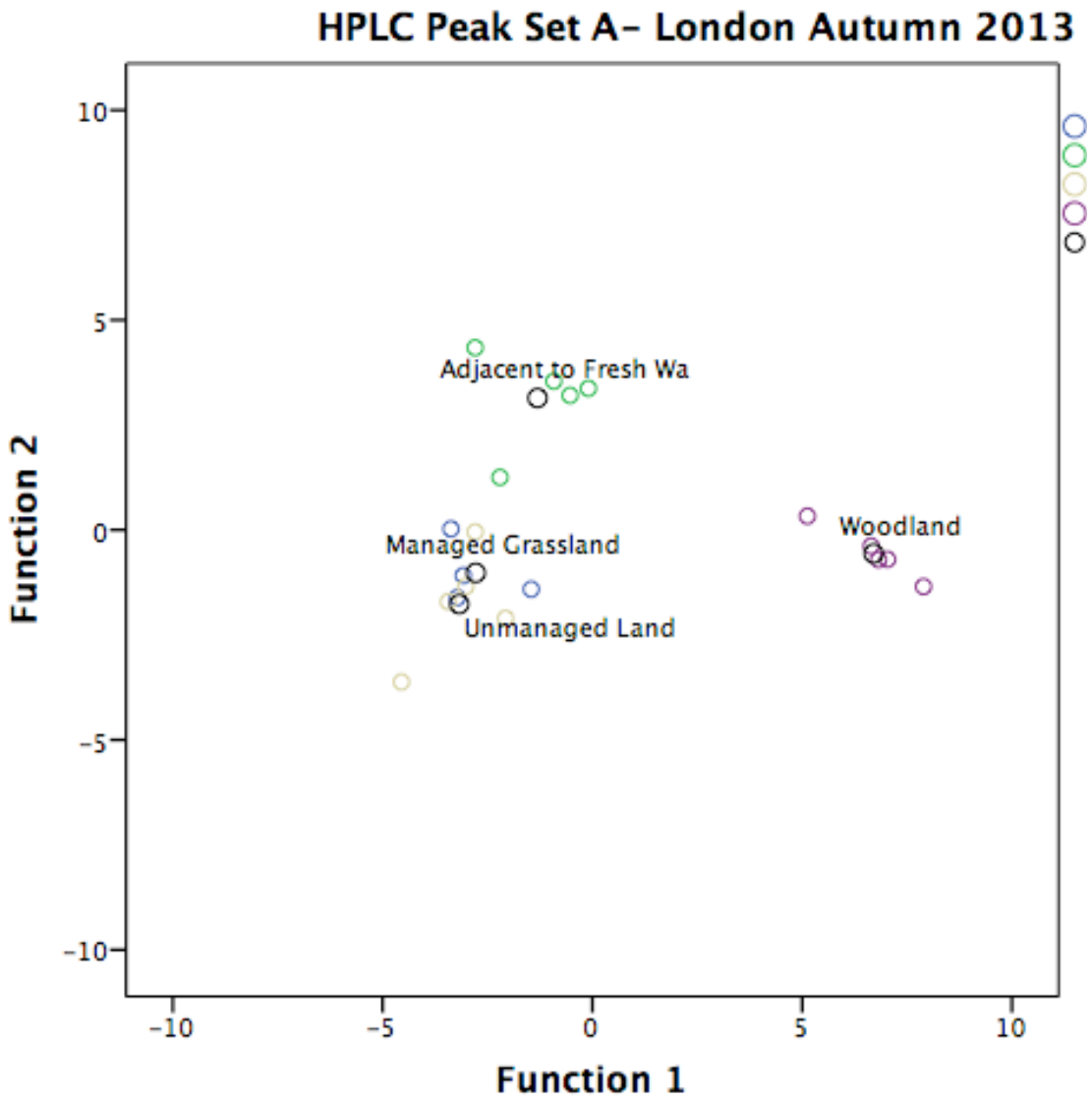
the same peak ratios were 2.1:1 and 1.6:1, respectively. The profiles or the soils from the locations adjacent to fresh water and managed grassland were visually very similar, and the most clearly visible difference in the profiles was the increase in height between the peaks at 13.6min and 15.3min, which was 49% for the managed grassland but only 16% for the location adjacent to fresh water.

Figure 6.35 Seasonal Changes to Peak Set A Profiles- London, Autumn 2013



The discrimination produced with the peak set A data (Figure 6.36, Table 6.2) was 89.5% accurate in grouping the samples collected in Autumn 2013 from the London site. The functions explained 80.0%, 18.0% and 2.0% of the differences between the groupings produced, and this discrimination was statistically significant at the 99% confidence interval (p=0.009). Using this data set two samples were misclassified, one sample that was originally from managed grassland was misclassified as having originated from unmanaged land and vice versa.

Figure 6.36 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set A- London, Autumn 2013

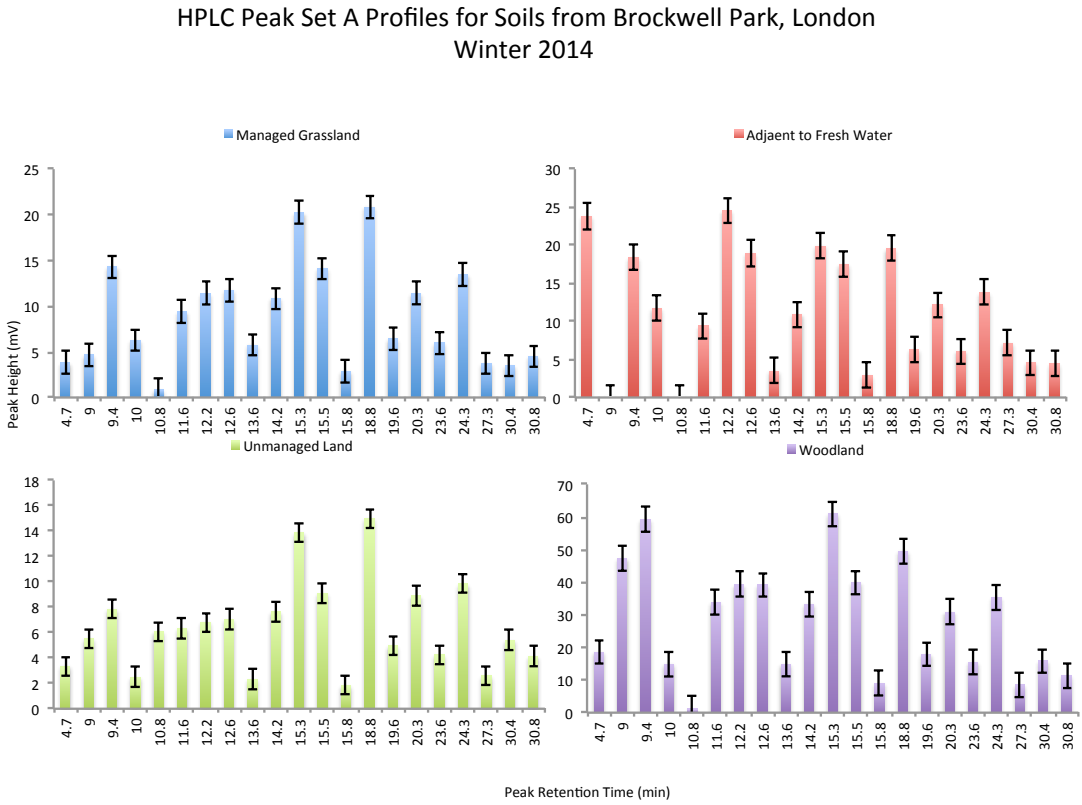


Winter 2014

All four locations within Brockwell Park, London could be distinguished by the profiles of HPLC peak set A (Figure 6.37), as discussed in Chapter 5. Samples adjacent to fresh water were distinct since the size order for the peaks at 4.7, 9.4 and 10min was different from the other three locations, and due to the absence of the peaks at 9 and 10.8min. Discrimination of managed grassland from woodland and unmanaged land was achieved through comparison of the size of the peak at 9.4min, since the ratio of this peak compared with the peak at 9min was larger for managed grassland, at 3:1, than for both unmanaged land and woodland, at 1.3:1 and 1.4:1, respectively. Woodland soil profiles differed from the other locations as the magnitudes of the peaks were more than double those of the other sites, while the relatively high ratio of 2.5:1 for the peak at 15.3min, relative to the peak at 9min, for woodland samples

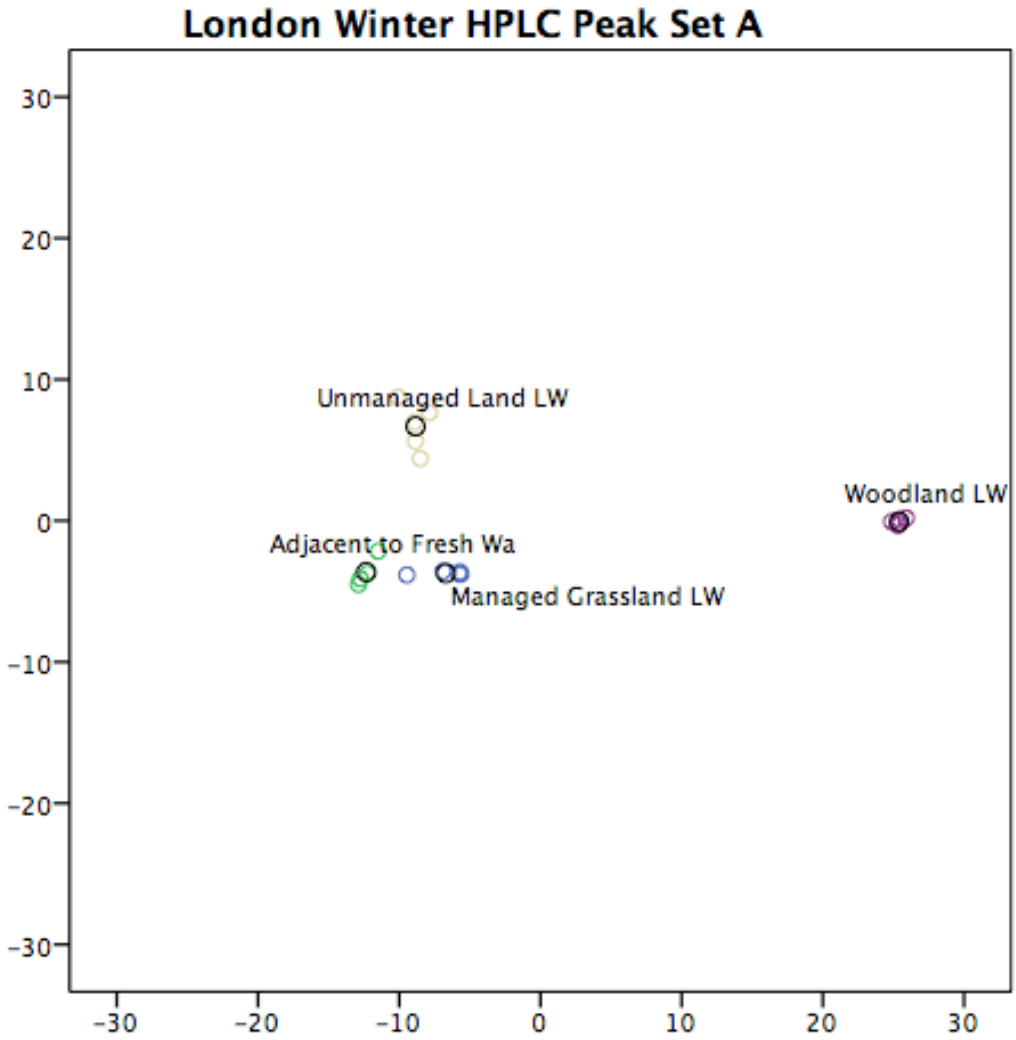
separated these samples from unmanaged land where the ratio was 1.2:1. Furthermore the two tallest peaks were 4.7 and 12.2min for soils adjacent to fresh water, whereas the largest peaks were at 9.4 and 15.3min for woodland soils.

Figure 6.37 Seasonal Changes to Peak Set A Profiles- London, Winter



100% accuracy in grouping samples from London in Winter was achieved using peak set A in the CDFA (Figure 6.38, Table 6.2). This discrimination was statistically significant at the 99% confidence interval (p=0.000) and the three functions produced explained 89.7%, 7.0% and 3.3% of the resulting variation in the samples.

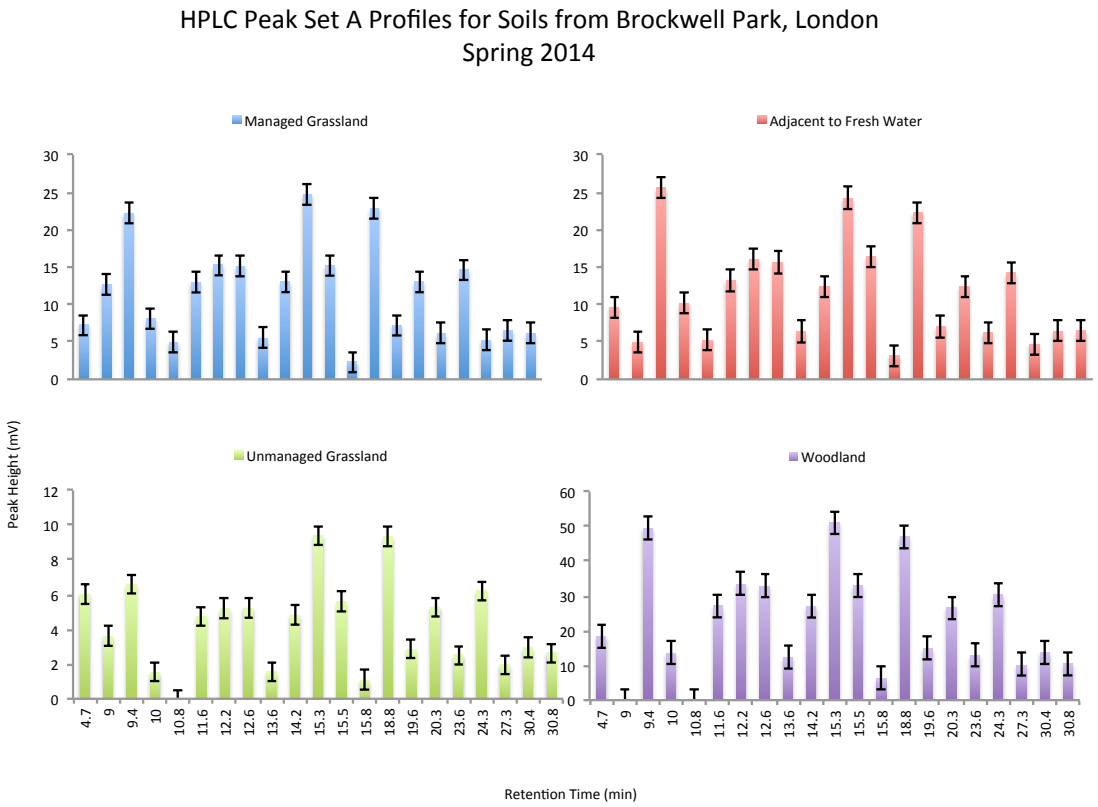
Figure 6.38 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set A- London, Winter



Spring 2014

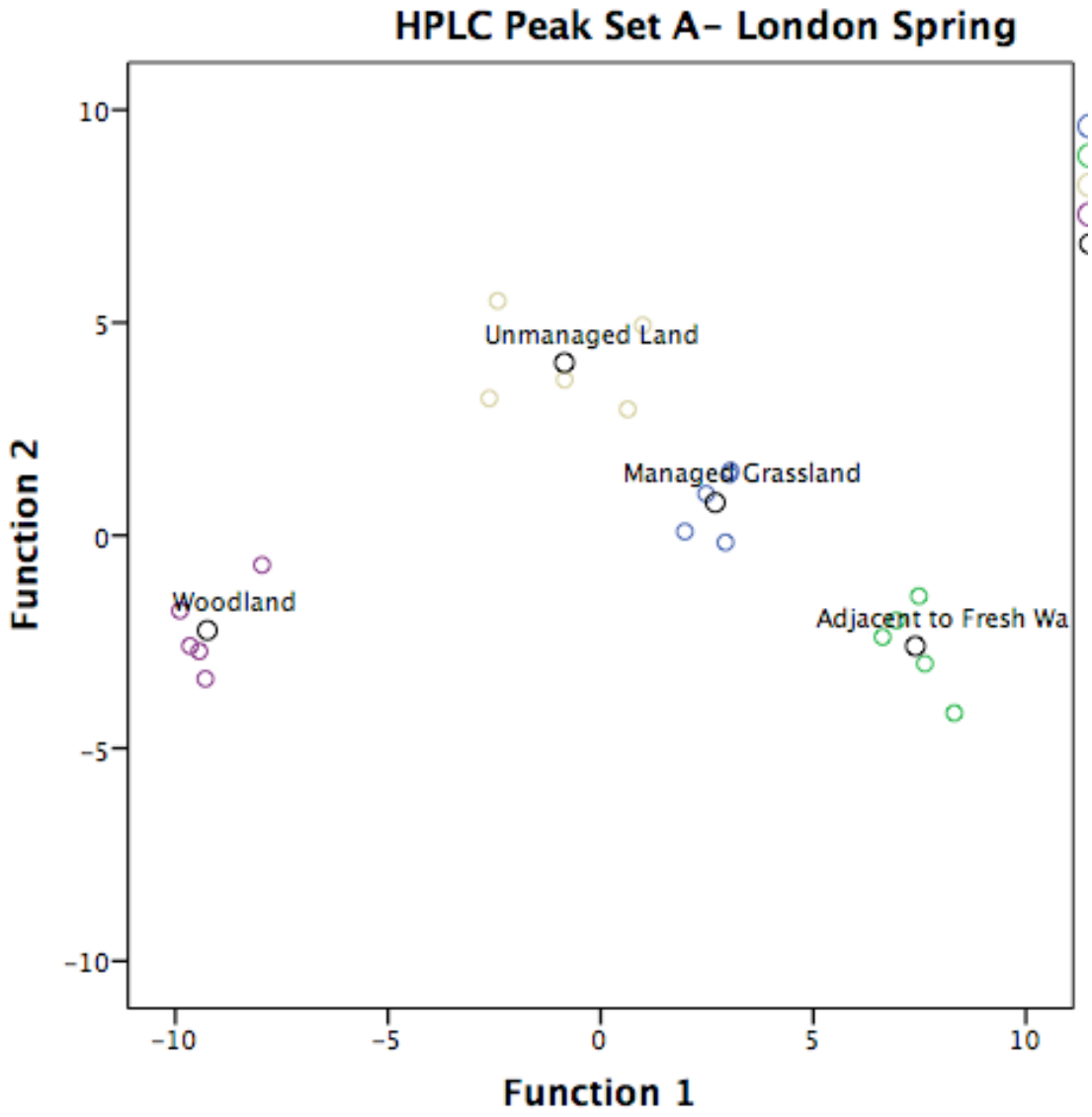
The peak set A profiles (Figure 6.39) for the soils from woodland could be discriminated from the other locations at the Spring 2014 time point through the absence of the peak at 10.8min in addition to the absence of the peak at 9min, while the presence of the peak 9min, in combination with the absence of the peak at 10.8min, was indicative of unmanaged soils. The remaining locations, from the soils adjacent to fresh water and managed grassland, could be discriminated by the relative heights of the peaks at 9.4min compared to the peak at 9min, as the peak at 9.4min was far larger than the peak at 9min for the location adjacent to fresh water, where the peak height ratio was 5.2:1, whereas for the managed grassland the ratio was 1.8:1.

Figure 6.39 Seasonal Changes to Peak Set A Profiles- London, Spring



The functions produced in the CDFA for samples from London in Spring (Figure 6.40, Table 6.2) explained 81.2%, 15.8% and 3.0% of the resulting variation in the samples, producing a discrimination that was 100% accurate and statistically significant at the 99% confidence interval (p=0.000)

Figure 6.40 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set A- London, Spring

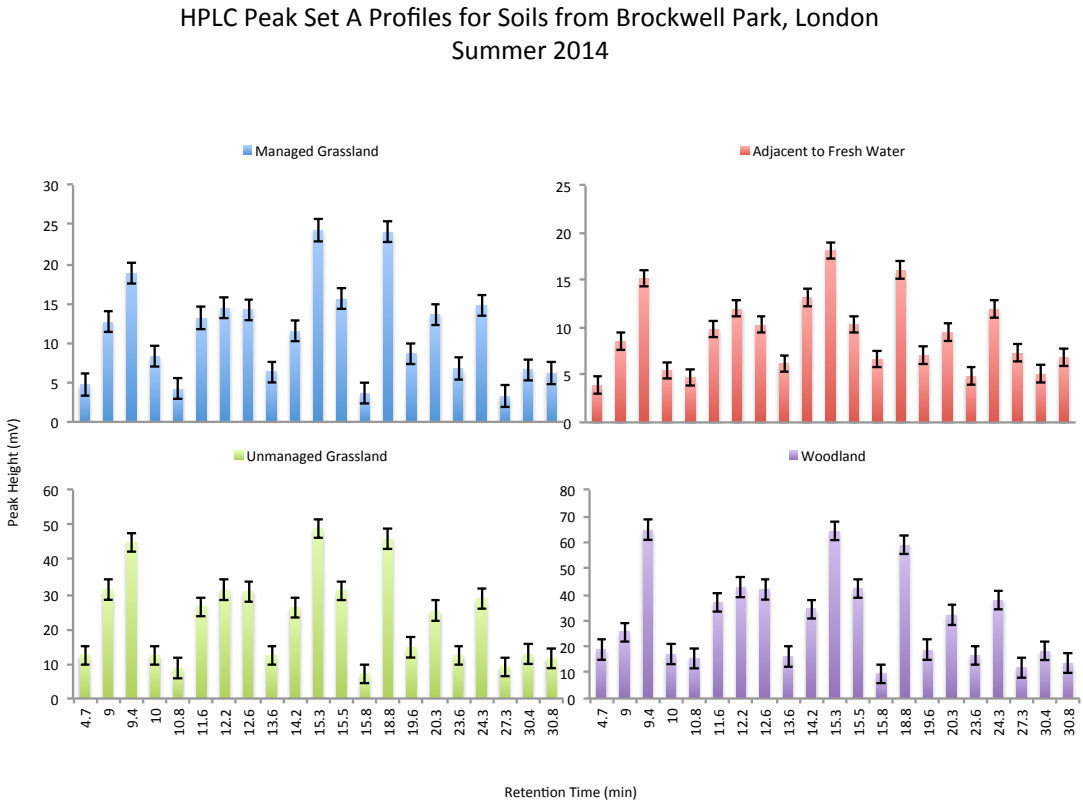


Summer 2014

At the Summer 2014 time point (Figure 6.41), the profiles of the soils from the woodland location could be discriminated from all others by the prominence of the peak at 9.4min, which was the largest peak for this location but not the others, and the difference in height between the peaks at 4.7min and 9min, the peak at 9min was only 38% larger than the peak at 4.7min, while the size increase was 118% for the location adjacent to fresh water, 170% for the managed grassland, and 149% for the unmanaged land. The ratio of the heights of the peaks at 15.5min compared to the peaks at 15.8min allowed the locations adjacent to fresh water to be excluded from the remaining two locations, as the ratio was only 1.6:1 compared to 4.4:1

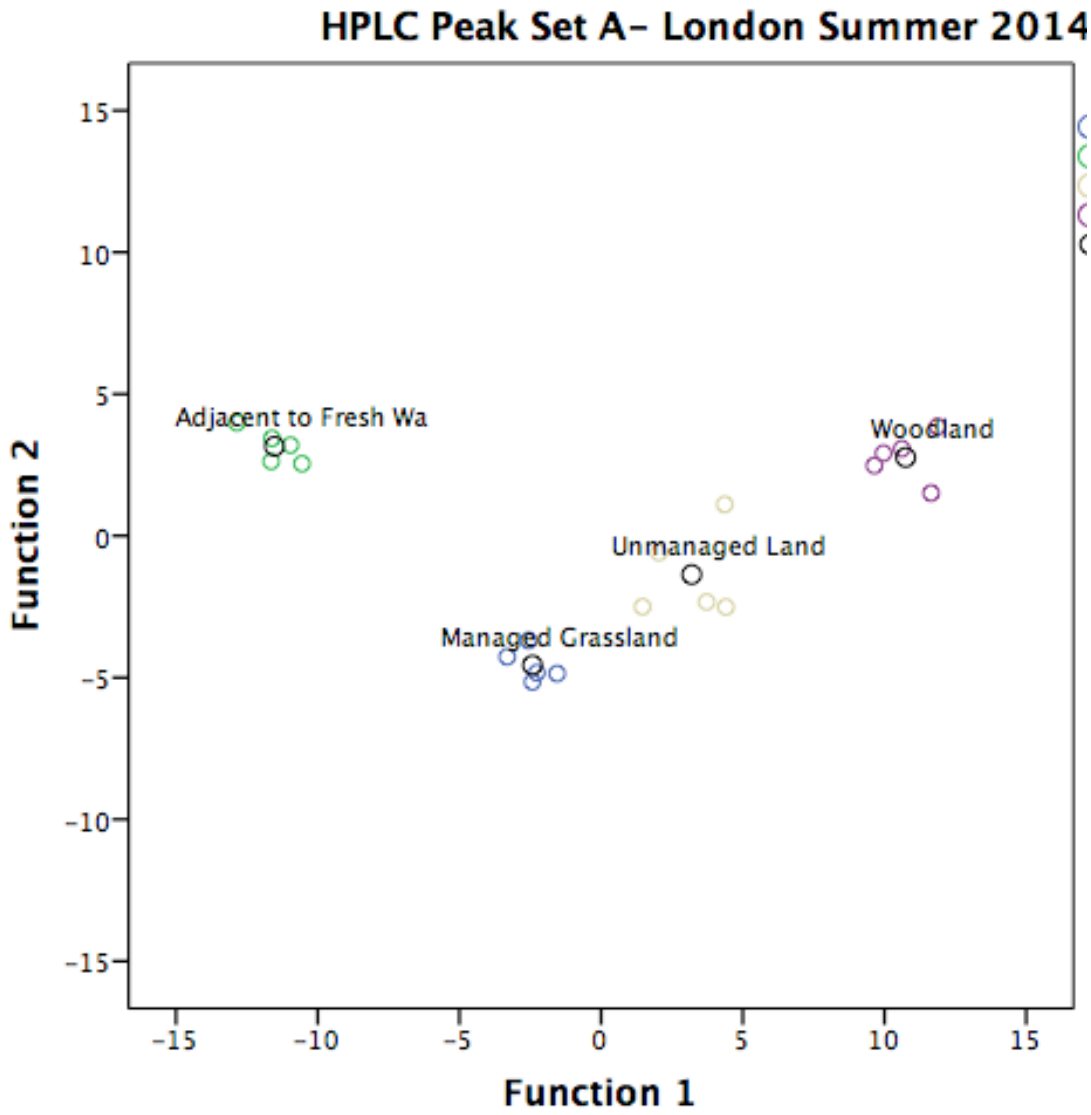
for both the managed grassland and unmanaged land. The profiles for the managed grassland and unmanaged land were the most visually similar, however these locations could be discriminated by the differences in the absolute peak heights, which ranged from 3.25mV to 24.07mV for the managed grassland, but were far larger for the unmanaged land, where the peak heights were in the range of 7.07mV to 48.91mV. In addition, the relative heights of the peaks at 10min and 10.8min allowed the two profiles to be discriminated, as the earlier peak was approximately twice the size of the peak at 10.8min for managed grassland, with peak heights of 8.28mV and 4.17mV, however for the unmanaged location, the peak at 10min was only 43% larger, at 12.41mV, than the peak at 10.8min, which was 8.67mV in height.

Figure 6.41 Seasonal Changes to Peak Set A Profiles- London, Summer 2014



For the Summer 2014 time point, the London data for peak set A gave 100% accuracy in the CDFA (Figure 6.42, Table 6.2), and this discrimination was statistically significant at the 99% confidence interval (p=0.000). The three canonical functions produced in this analysis explained 83.9%, 12.8%, and 3.3% of the variation between the samples.

Figure 6.42 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set A- London, Summer 2014

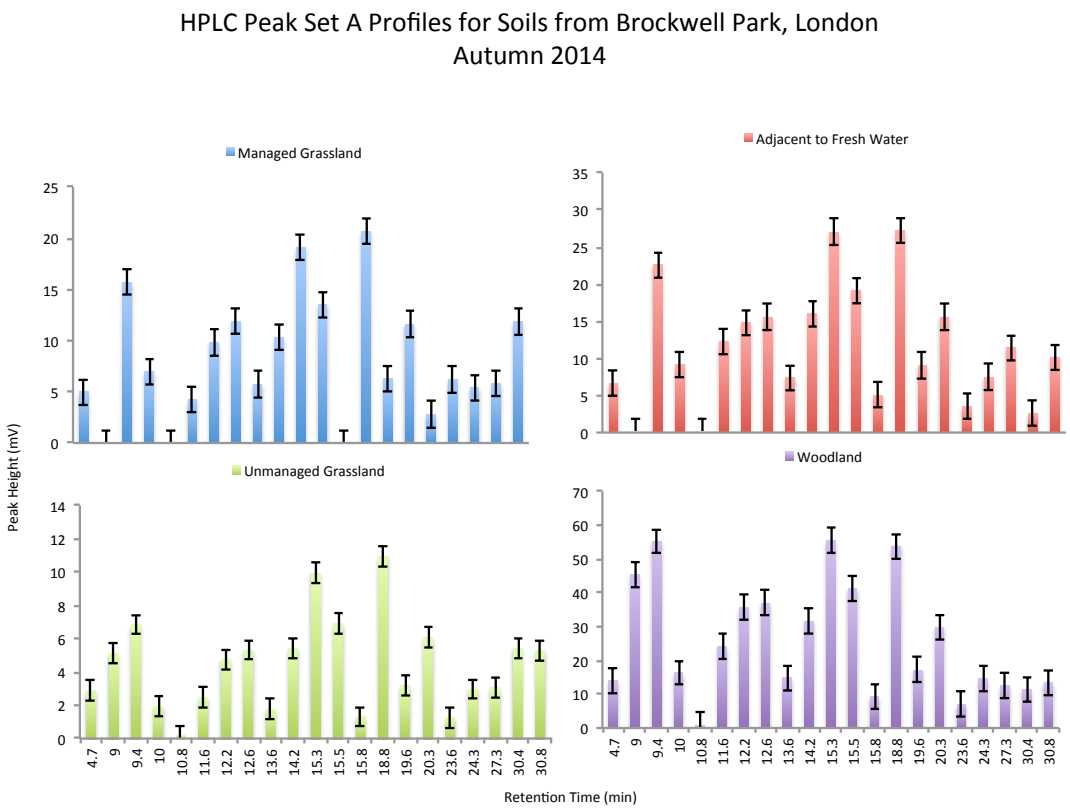


Autumn 2014

The profiles of the location adjacent to fresh water were easy to identify through the absence of peaks at 9min and 10.8min, while the managed grassland profiles could be identified by the absence of these two peaks in addition to the peak at 15.8min (Figure 6.43). All the peaks in peak set A were present in the profiles of the soils from the woodland and unmanaged locations, however these two locations could be visually discriminated on the basis of the relative heights of the peaks at 9.4min and 15.3min, since the later peak was 45% larger at 9.93mV for the unmanaged land but the two peaks were approximately equal at 55.29mV and 55.36mV in height, respectively, for the woodland location. The woodland profiles could be further discriminated through comparison of the absolute peak heights, since the maximum

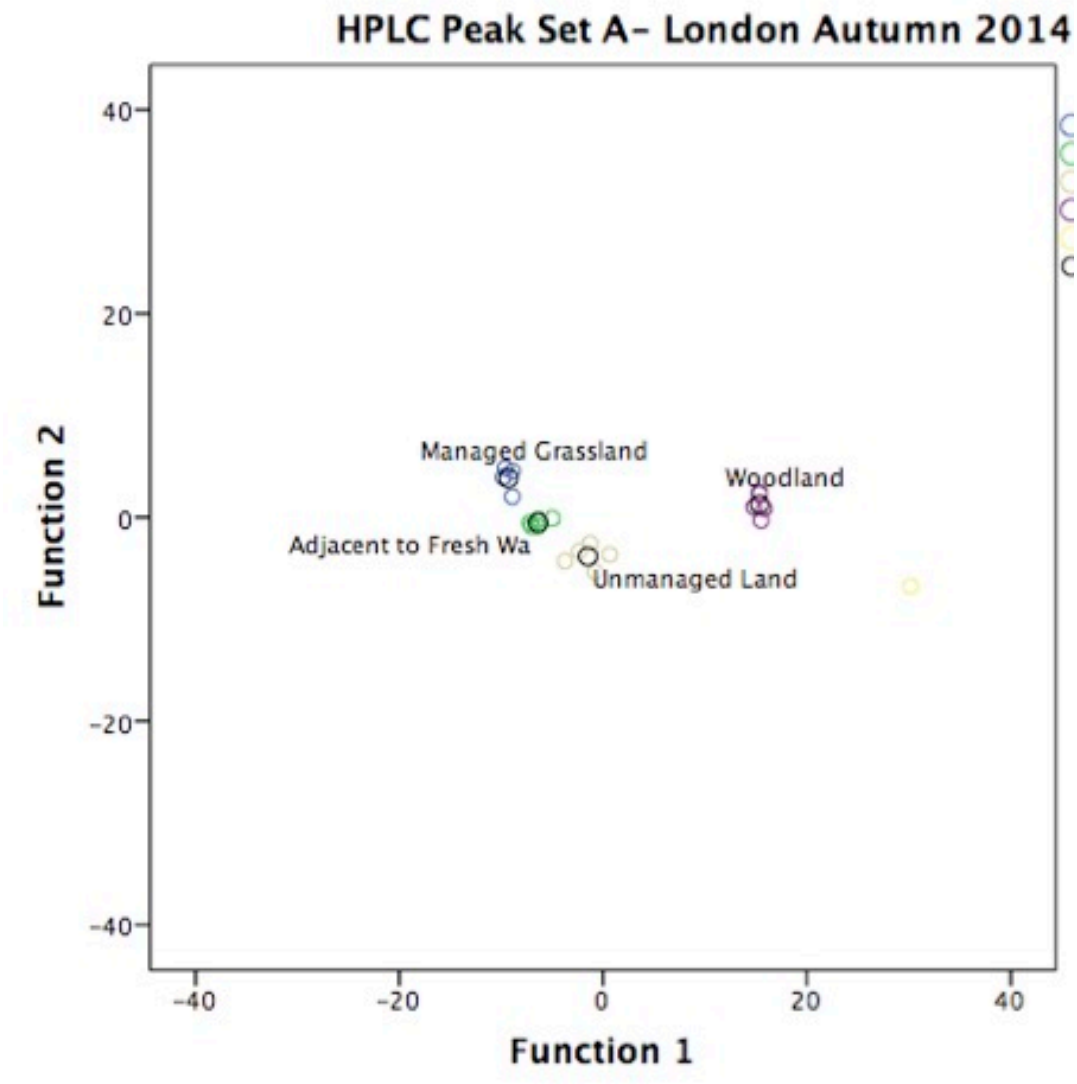
peak height was 55.36mV at 15.3min, while the maximum peak height was only 27.3mV across the three other sample locations

Figure 6.43 Seasonal Changes to Peak Set A Profiles- London, Autumn 2014



In Autumn 2014, the CDFA for peak set A (Figure 6.44, Table 6.2) afforded 100% accuracy in grouping the London samples, with 91.4%, 7.4% and 1.1% of the variability in the samples explained by the first three functions. This discrimination was statistically significant at the 99% confidence interval (p=0.000)

Figure 6.44 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set A- London, Autumn 2014



6.3.3.2 HPLC Profiles for Peak Set B

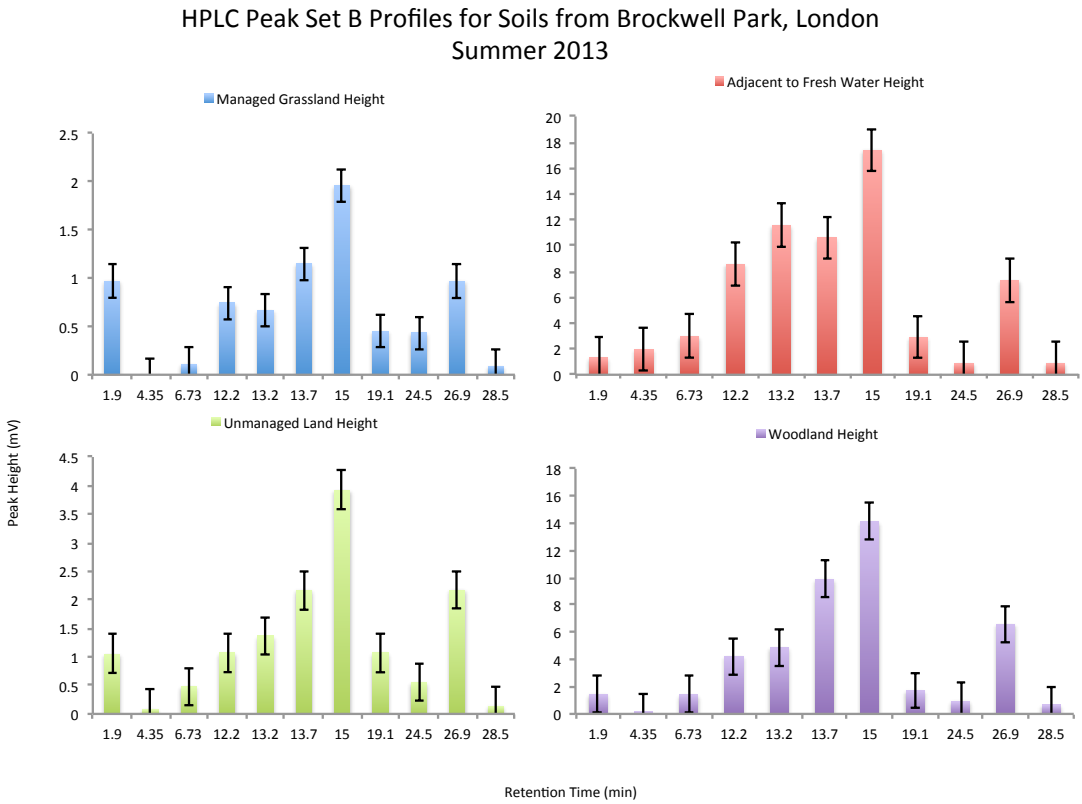
The profiles obtained using peak set B were useful for distinguishing the four locations within Brockwell Park, London, and the primary descriptors used to visually classify the soil profiles according to location are summarised below.

Summer 2013

The peak set B profiles for the London site at the Summer 2013 time point (Figure 6.45) could be separated by visual comparison of the relative sizes of the peaks at 1.9min and 12.2min, since the earlier peaks were 66% and 84% smaller, respectively, than the later peak for soils adjacent to fresh water and woodland soils, while for managed grassland the earlier peak was 29% larger and for unmanaged land the two peaks were equal in size, as the earlier peak was only 1.9% smaller. Managed grassland profiles were further separated from the unmanaged

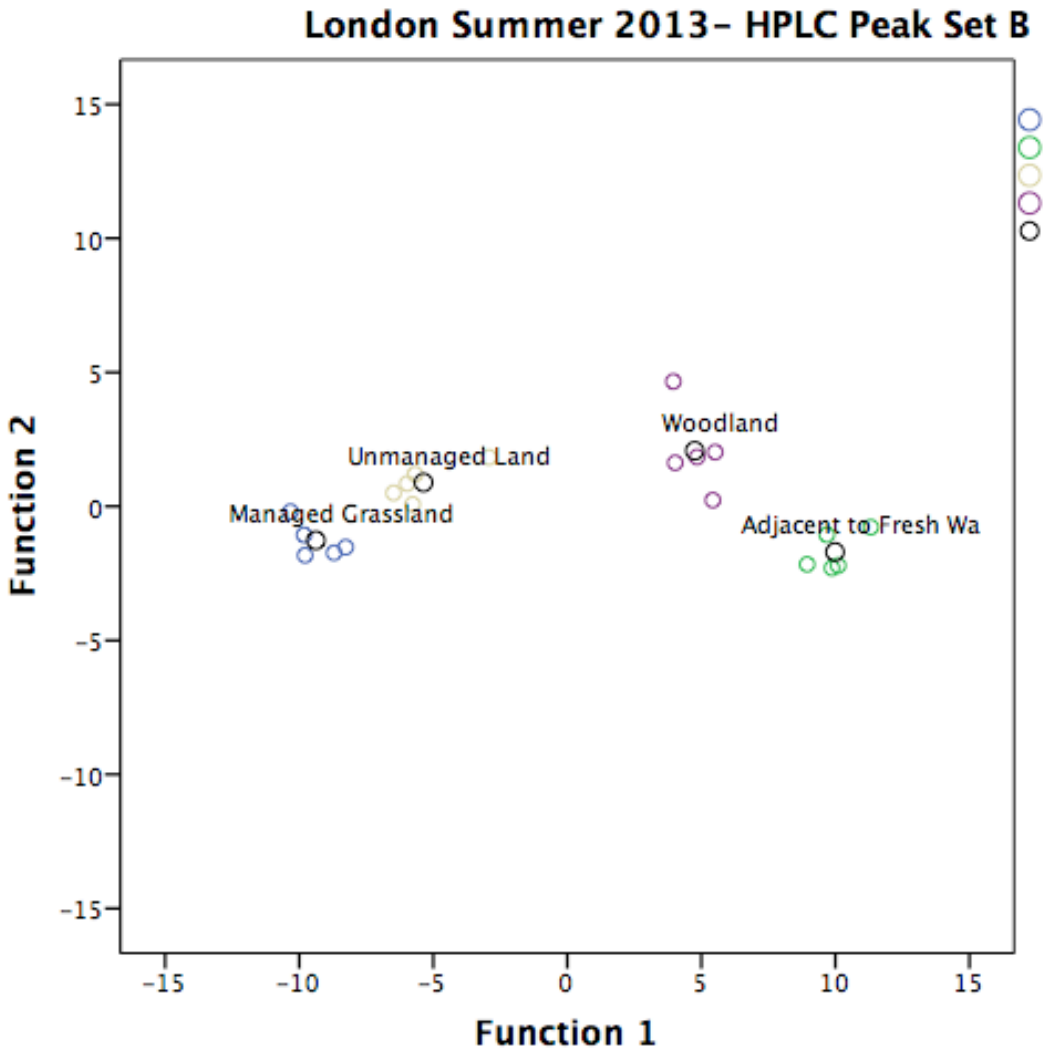
location by the relative heights of the peaks at 1.9min and 26.9min, which were equal at 0.97mV for the former location, however for the latter location the peak at 26.9min was more than twice the size of the peak at 1.9min, at 1.05mV and 2.16mV respectively. The woodland locations and location adjacent to fresh water could be further separated from one another through the presence of a peak at 4.35min for the latter location that was absent at the former, and the ratio of the peak height at 13.7min compared to 13.2min, which was 0.9:1 for locations adjacent to fresh water and 2:1 for woodland locations.

Figure 6.45 Seasonal Changes to Peak Set B Profiles- London, Summer 2013



100% Accuracy in grouping the samples from London in Summer 2013 was achieved when peak set B was used in the CDFA (Figure 6.46, Table 6.3), and this discrimination was significant at the 99% confidence interval (p=0.000). The three discriminant functions produced in this analysis explained 95.2%, 3.8% and 0.9% of the variation in the samples.

Figure 6.46 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set B- London, Summer 2013

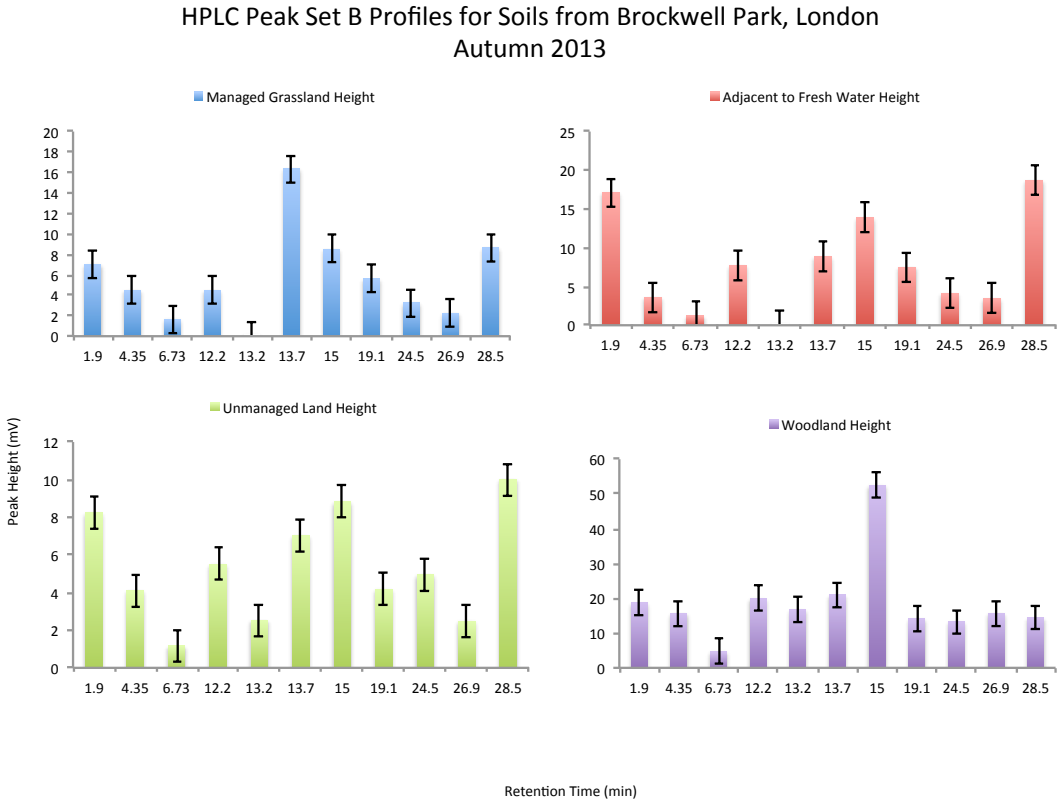


Autumn 2013

At the Autumn 2013 time point, the peak set B profiles (Figure 6.47) could be separated into two groups based on the presence of a peak at 13.2min for woodland and unmanaged land, which was absent at the other locations. The woodland profiles could be visually separated from the unmanaged land due to the prominent peak at 15min, which was 150% larger for the woodland soils than the next largest peak at 13.7min, whereas for unmanaged land the peak at 15min was 16% smaller than the largest peak at 28.5min, and was only 25% larger than the peak at 13.7min. Likewise, the managed grassland and location adjacent to fresh water could be discriminated by the large prominent peak at 13.7min in the managed grassland profiles which, at 16.38mV, was 89% larger than the next largest peak at 28.5min, in comparison the peak at 28.5min was the largest peak in the profiles of the soils from locations adjacent to

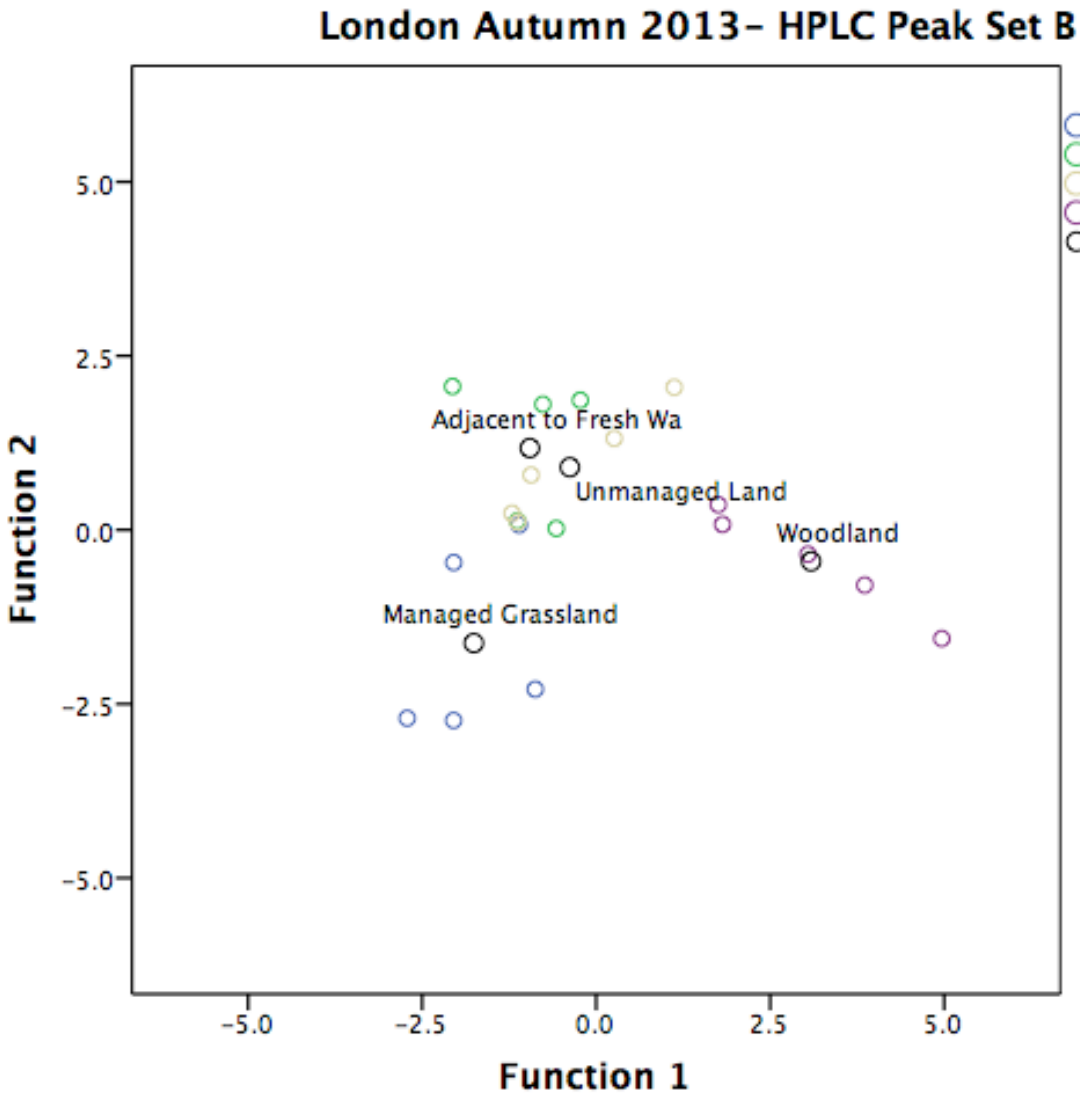
fresh water and the peak at 13.7min was 53% smaller, with peak heights of 18.74mV and 8.82mV, respectively.

Figure 6.47 Seasonal Changes to Peak Set B Profiles- London, Autumn 2013



The discrimination produced with the peak set B data was 80% accurate in grouping the samples collected in Autumn 2013 from the London site (Figure 6.48 Table 6.3), as one sample from the managed grassland location was misclassified as originating from unmanaged land, one sample from unmanaged land was misclassified as originating from the location adjacent to fresh water, and conversely two of the samples from the location adjacent to fresh water were misclassified as having originated from unmanaged land using the functions generated. The functions explained 67.0%, 24.7% and 8.4% of the differences between the groupings produced, however this discrimination was not statistically significant (p=0.377)

Figure 6.48 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set B- London, Autumn 2013

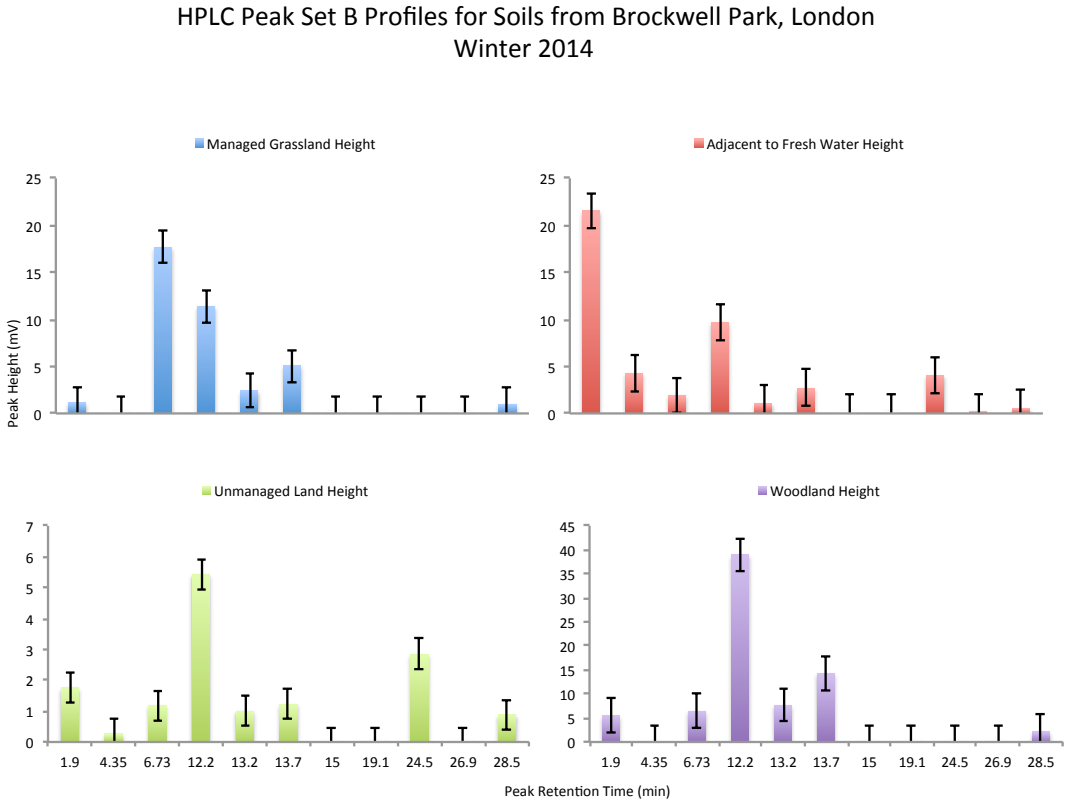


Winter 2014

The HPLC profiles for peak set B allowed all four samples locations within Brockwell Park, London to be discriminated visually (Figure 6.49), as presented in Chapter 5. The magnitudes of the peaks in woodland soils were approximately twice the size of those in soils adjacent to fresh water and managed grassland, with an average peak height of 39.0 mV for the largest peak at 12.2min for woodland soils, where the largest peaks, for soils adjacent to fresh water and managed grassland were 21.6 and 17.7mV respectively, while the peaks found in unmanaged land were around three times smaller at only 5.4mV. Managed grassland could be discriminated from all other sample locations through the large size of the peak at 6.73min relative to the peak at 12.2min, while the profiles in soils adjacent to fresh water differed from the other locations through the large size of the peaks at 1.9min compared to all other peaks.

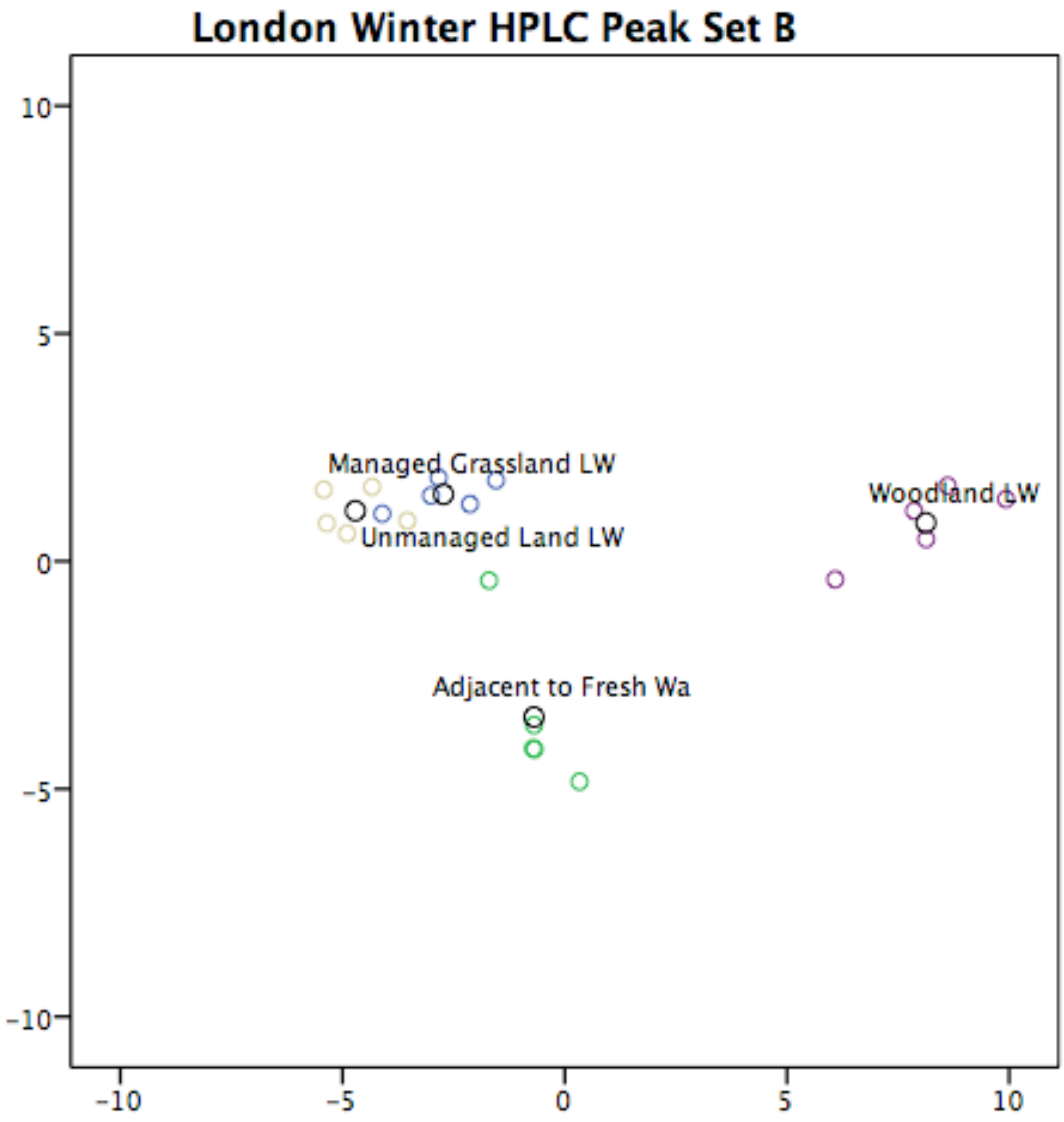
The profiles of soils from unmanaged land and woodland could be separated since the peak at 24.5min was not detected in the woodland samples.

Figure 6.49 Seasonal Changes to Peak Set B Profiles- London, Winter



In Winter, the CDFA for peak set B (Figure 6.50, Table 6.3) afforded 90% accuracy in grouping the London samples, with one sample from managed grassland misclassified as unmanaged land and one sample from the location adjacent to fresh water misclassified as having originated from managed grassland. The functions generated in the analysis explained 84.8%, 13.9% and 1.3% of the variability in the samples, and this discrimination was statistically significant at the 99% confidence interval (p=0.000).

Figure 6.50 Plot of sample scores for Function 2 (y-axis) vs. sample scores for Function 1 (x-axis) for Peak Set B- London, Winter

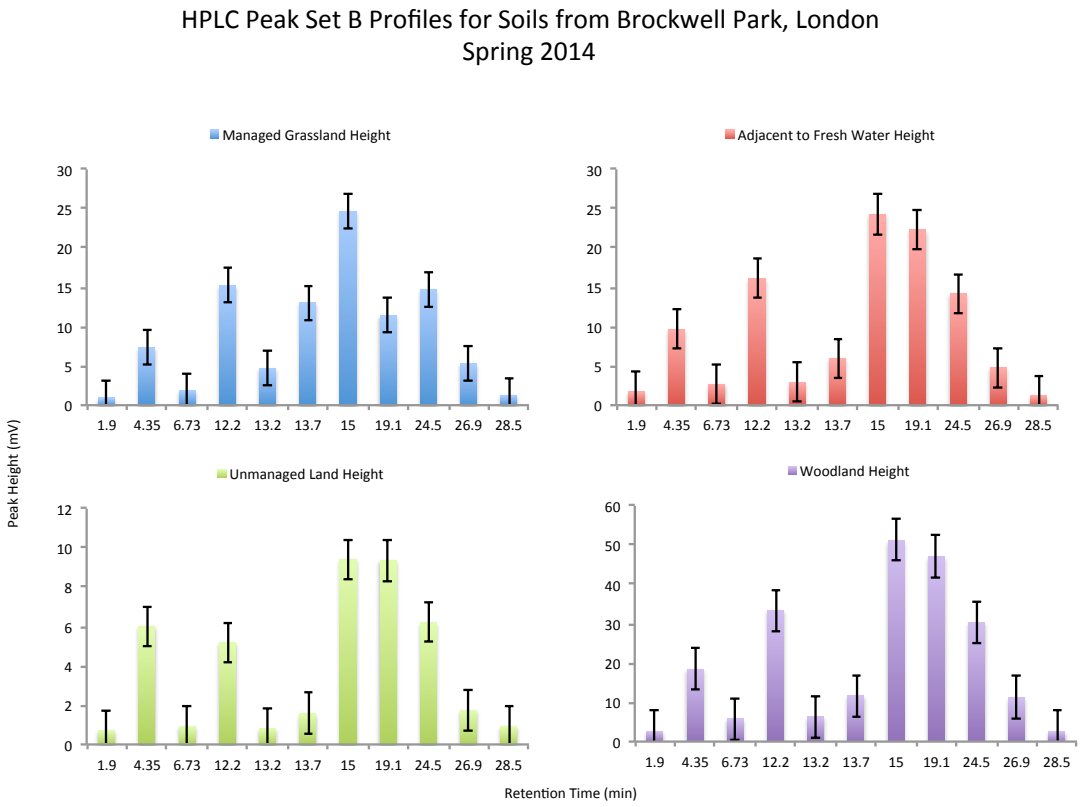


Spring 2014

At the Spring time point, managed grassland profiles (Figure 6.51) were unambiguously identifiable visually by the presence of a single major peak at 15min, which was 63% larger than the next largest peak at 12.2min, whereas at all other locations there were two prominent peaks at 15min and 19.1min that were approximately equal in size, with the later peaks 7.8%, 0.8%, and 8.1% smaller in height for the location adjacent to fresh water, unmanaged land and woodland, respectively. The profiles of unmanaged land could be excluded from the remaining two locations since the peak at 4.35min was 16% larger than the peak at 12.2min, whereas the peak at 4.35min was 40% and 45% smaller than the peak at 12.2min for the profiles of the location adjacent to fresh water and woodland, respectively.

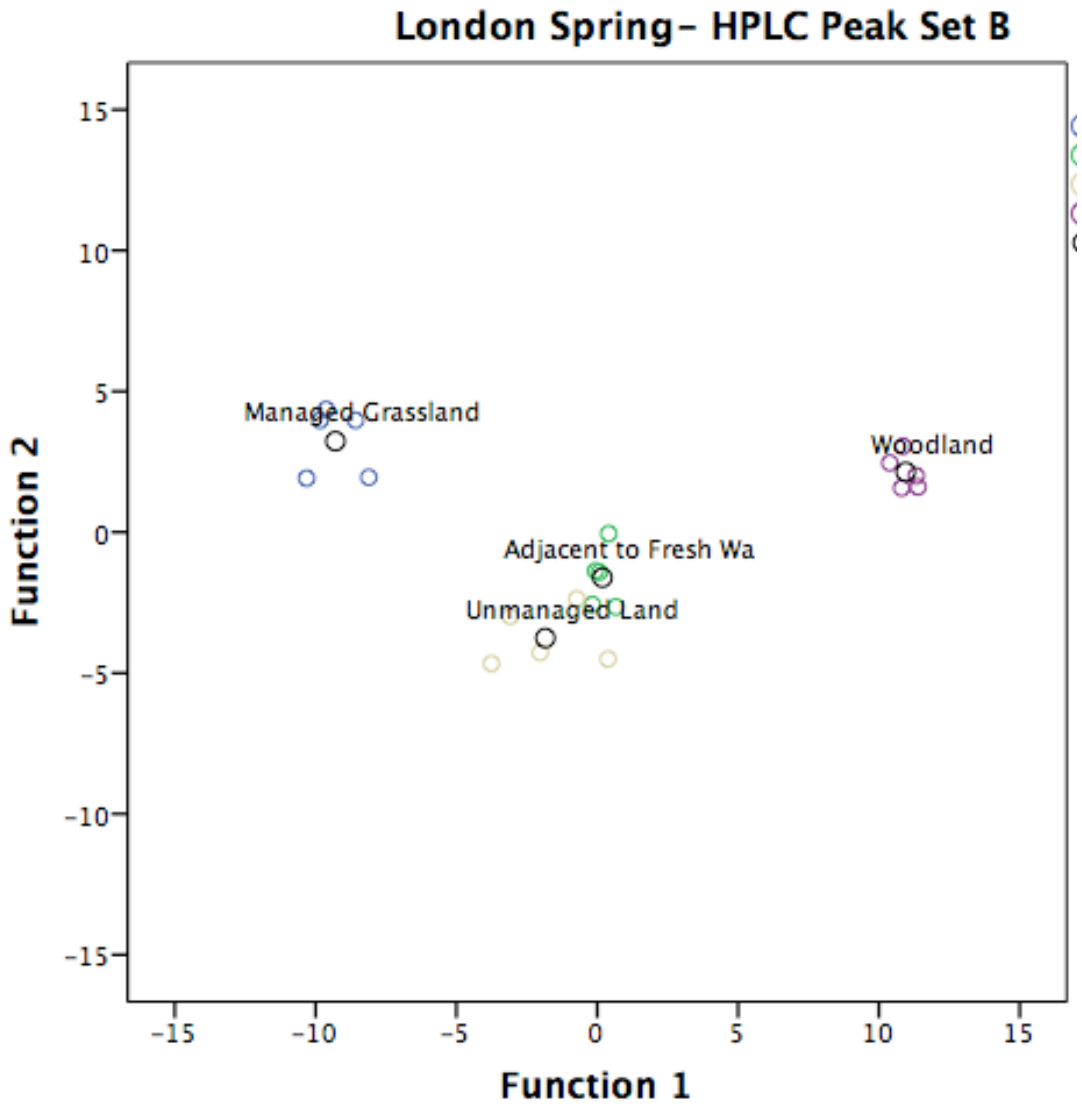
There were no clear visual differences in the relative peak sizes in the profiles obtained for the location adjacent to fresh water and woodland, however the absolute peak heights for woodland soils were in the range of 2.54-51.2mV; more than double the size of the peaks in the profiles of soils from locations adjacent to fresh water, which ranged from 1.22-24.22mV.

Figure 6.51 Seasonal Changes to Peak Set B Profiles- London, Spring



For the Spring time point, the London data for peak set B gave 100% accuracy in the CDFA (Figure 6.52, Table 6.3), and this discrimination was statistically significant at the 99% confidence interval (p=0.000). The three canonical functions produced in this analysis explained 86.4%, 13.1%, and 0.5% of the variation between the samples.

Figure 6.52 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set B- London, Spring

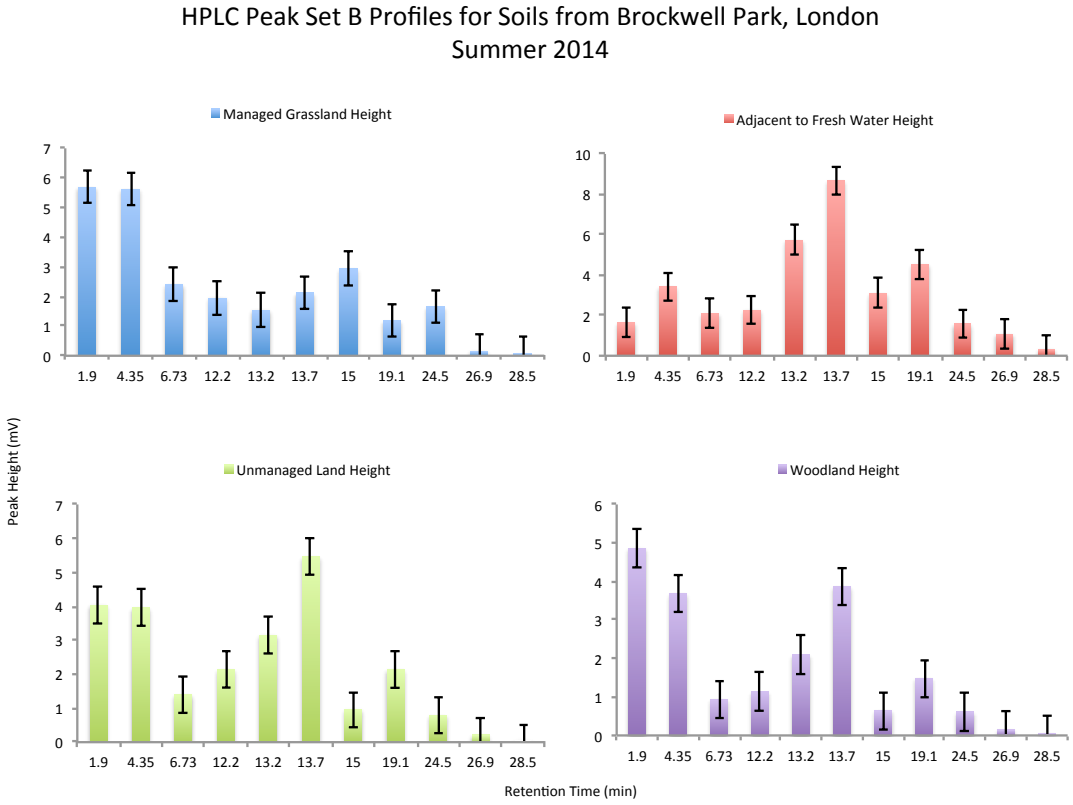


Summer 2014

The profiles obtained for peak set B at the Brockwell park site in Summer 2014 (Figure 6.53) could be divided into two groups by visual comparison of the relative height of the peaks at 1.9min and 4.35min, which were approximately equally sized for both managed grassland and unmanaged land, as the later peak was only 1.2% and 1.7% larger, respectively, for these locations. Comparison of the retention times of the largest peak allowed these two locations to be discriminated, as the two peaks at 1.9min and 4.35min were the largest in the profiles of the managed grassland soils but the peak at 13.7min was 36% greater in height than these peaks for the profiles from unmanaged land.

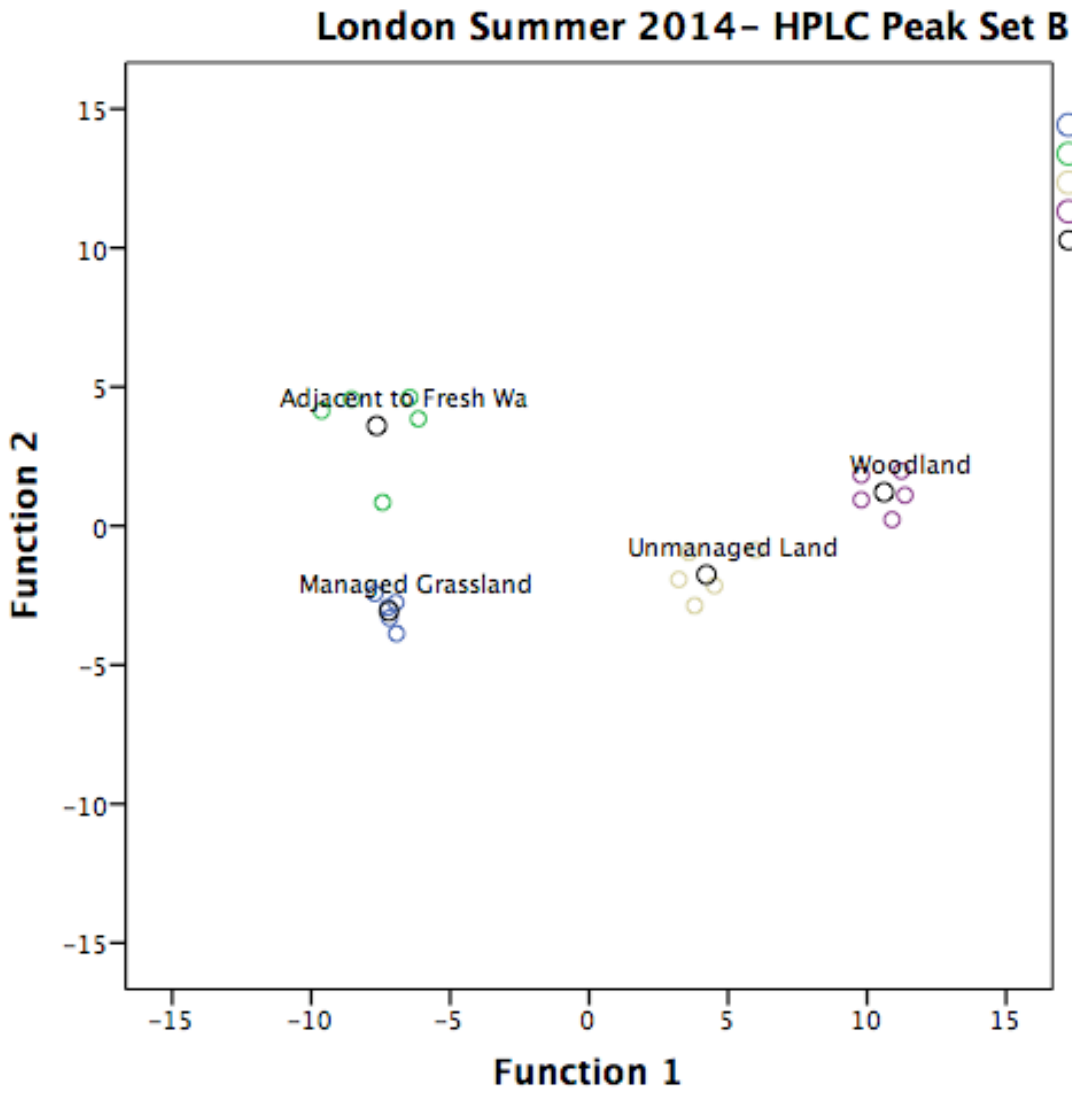
In contrast, the peak at 4.35min was 52% smaller than the peak at 1.9min for the woodland soil profiles, while for the location adjacent to fresh water the same peak was 32% larger, and the two profiles could also be discriminated since the peak at 13.7min was the largest in the profile obtained for the location adjacent to fresh water, while it was the second largest peak in the profiles obtained at the woodland location.

Figure 6.53 Seasonal Changes to Peak Set B Profiles- London, Summer 2014



100% accuracy in grouping samples from London in Summer 2014 was achieved using peak set B in the CDFA (Figure 6.54, Table 6.3). This discrimination was statistically significant at the 99% confidence interval (p=0.000) and the three functions produced explained 87.4%, 9.8% and 2.8% of the resulting variation in the samples.

Figure 6.54 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set B- London, Summer 2014

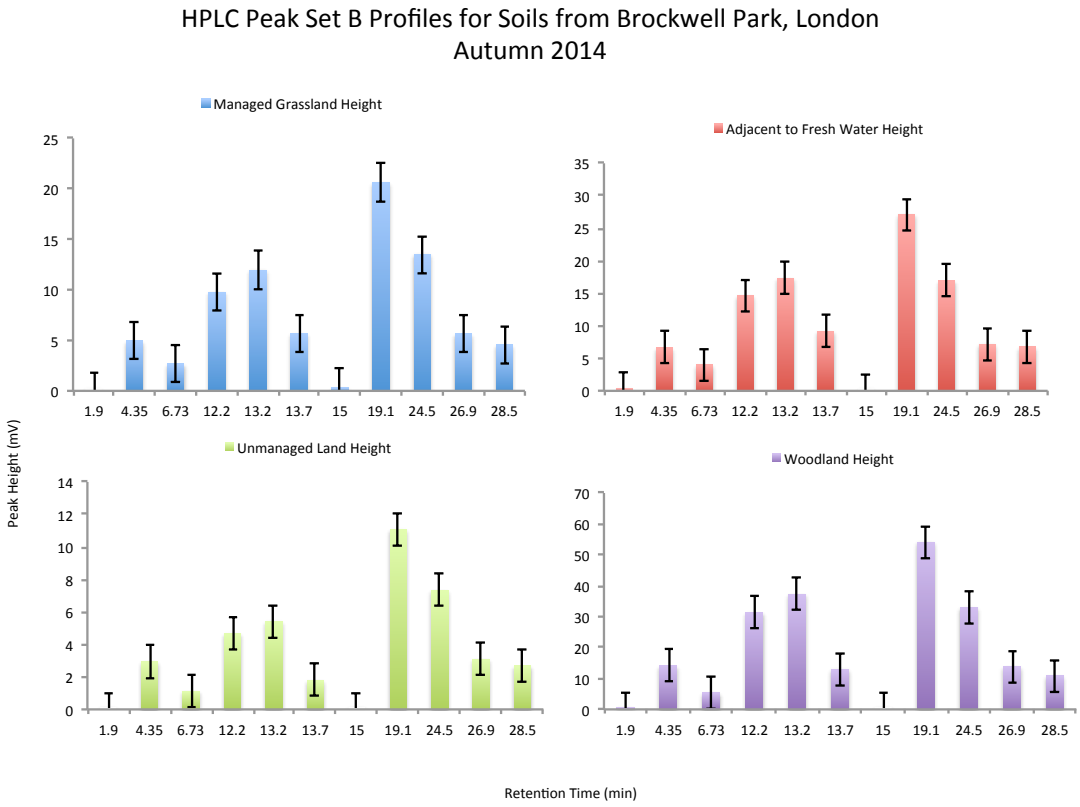


Autumn 2014

The overall profiles of each location at the London site at the Autumn 2014 time point were very similar when compared visually (Figure 6.55), as there were few differences in the relative peak heights between the locations. The woodland profiles could be discriminated from the other locations due to the large difference in absolute peak sizes, which were at least double the height of the other locations, with a maximum peak height of 54.1mV at 19.1min compared to 20.6mV, 27.1mV, and 11.1mV for the managed grassland, the location adjacent to fresh water, and unmanaged land. Unmanaged land could be separated from the remaining locations through comparison of the ratio height of the peaks at 13.2min to 13.7min, which was 3.0:1 for unmanaged land but 2.1:1 and 1.9:1, respectively for managed land and the location adjacent to fresh water. It was far more difficult to detect any visual differences

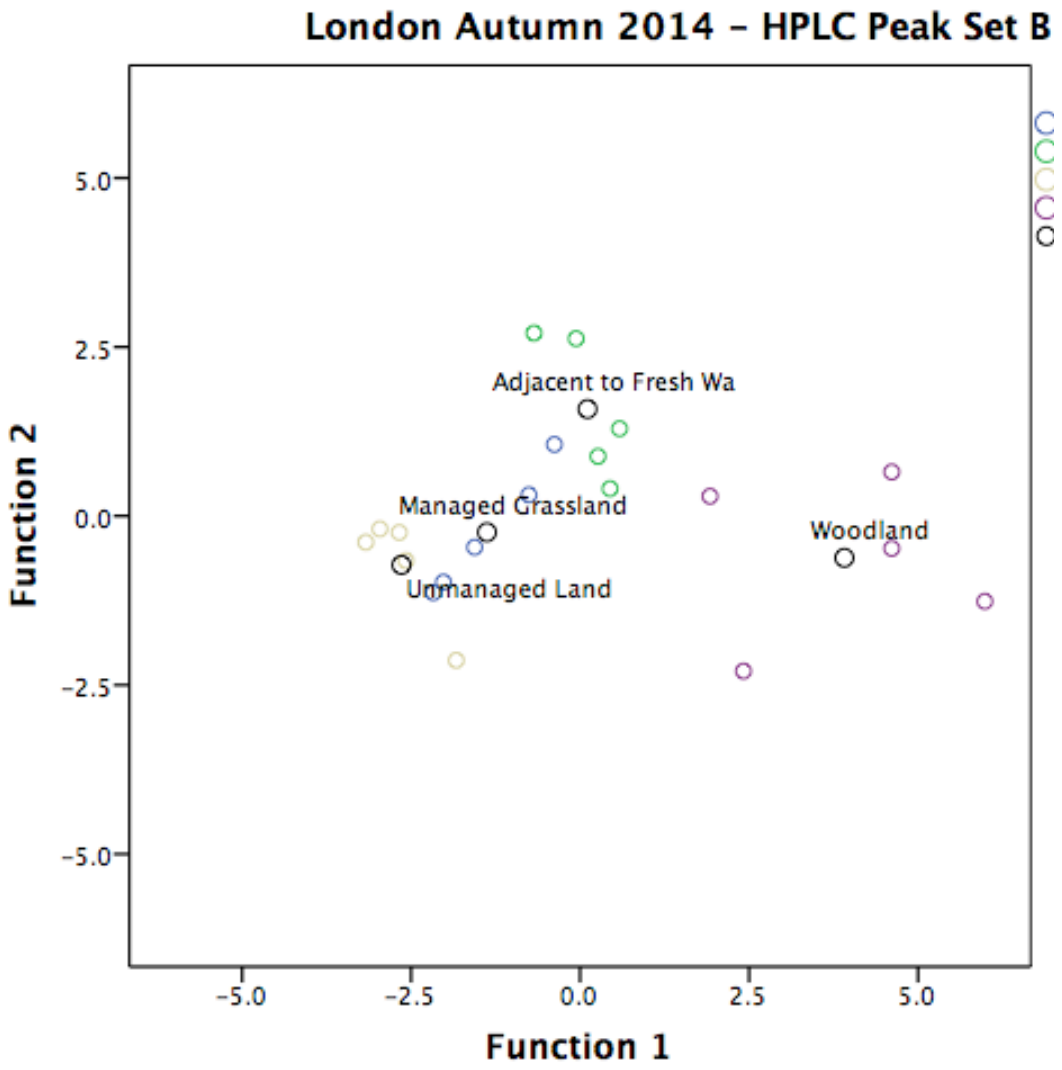
between the profiles for managed grassland and soils adjacent to fresh water, the only distinguishing features were the presence of a small peak of 0.34mV at 1.9min in the location adjacent to fresh water that was absent in the profiles from managed grassland, the presence of a small peak at 0.31mV at 15min for managed grassland that was absent from the profiles of the location adjacent to fresh water, and subtle differences in the ratio of the peaks at 19.1min and 28.5min, which were 4.5:1 and 4.0:1, respectively for the two locations.

Figure 6.55 Seasonal Changes to Peak Set B Profiles- London, Autumn 2014



The functions produced from peak set B in the CDFA for samples from London in Autumn 2014 (Figure 6.56, Table 6.3) explained 78.6%, 11.3% and 10.1% of the resulting variation in the samples, producing a discrimination that was 95% accurate. One managed grassland sample was misclassified as having originated from the location adjacent to fresh water, but this discrimination was not statistically significant (p=0.062)

Figure 6.56 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set B- London, Autumn 2014



6.3.4 New York City

Samples were collected from Central Park at the Winter time point only, and these results have been discussed in Chapter 5, however the results have been included in this chapter in order to aid the examination of the geographic variability of the profiles.

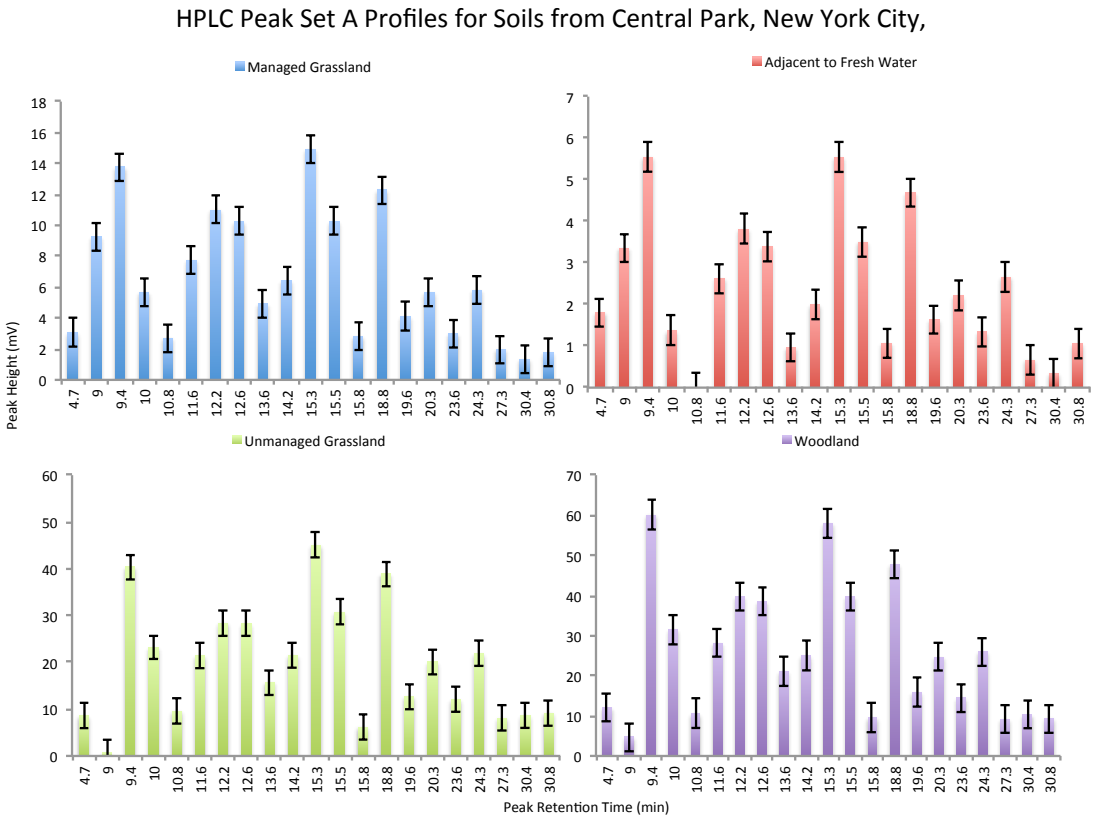
6.3.4.1 Peak Set A

The results of the visual comparison of the Peak Set A profiles obtained at Central Park, New York City, and the results of the CDFA using the peaks in set A as variables are detailed below

In Central Park, New York City, as was presented in Chapter 5, all four locations could be discriminated on the basis of their HPLC profiles for peak set A (Figure 6.57). The location adjacent to fresh water could be discriminated from all other locations through the absence of

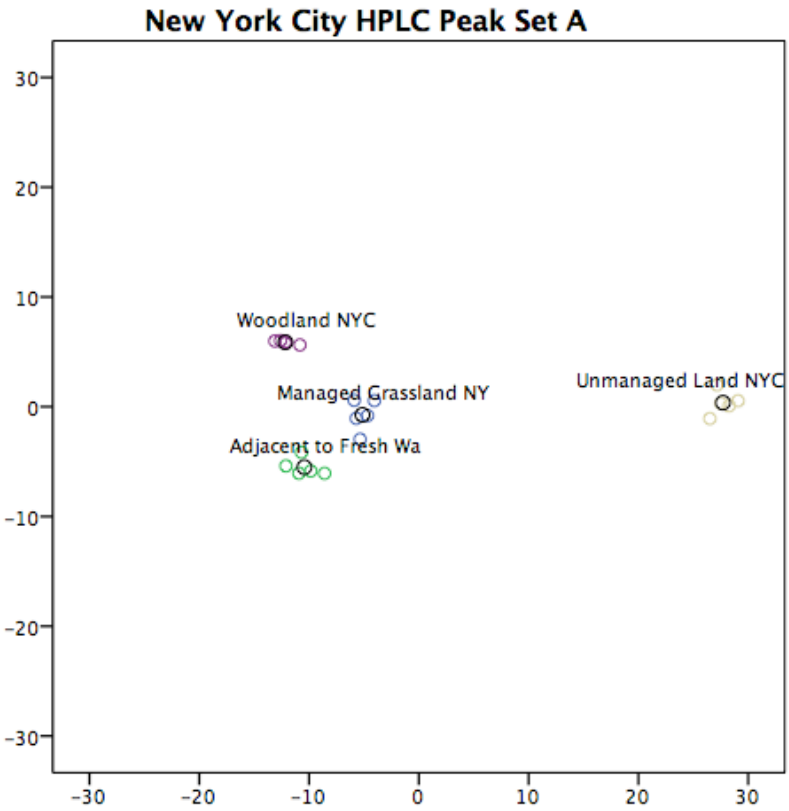
a peak at 10.8min, while for unmanaged land the absence of the peak at 9min was a unique feature among the locations at this site. The profiles of the managed grassland and woodland locations were different since the peaks were generally three times larger for the woodland soils, ranging from 5 to 60mV, than for managed grassland where the peaks ranged from 1 to 15mV, and the peak at 9min was larger, at 9mV, than the peak at 4.7min, at 3mV, for the managed grassland samples, while for the woodland location the peak at 4.7min was larger at 12mV than the peak at 9min, which was 5mV.

Figure 6.57: HPLC Profiles for Central Park, New York City- Peak Set A



Using peak set A, the CDFA classified the samples from New York City with 100.0% accuracy (Figure 6.58, Table 6.2), and this discrimination was statistically significant at the 99% confidence level (p=0.000). The first three functions explained 92.3%, 5.8%, and 1.9% of the variability in the sample groupings.

Figure 6.58 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set A- New York City



6.3.4.2 Peak Set B

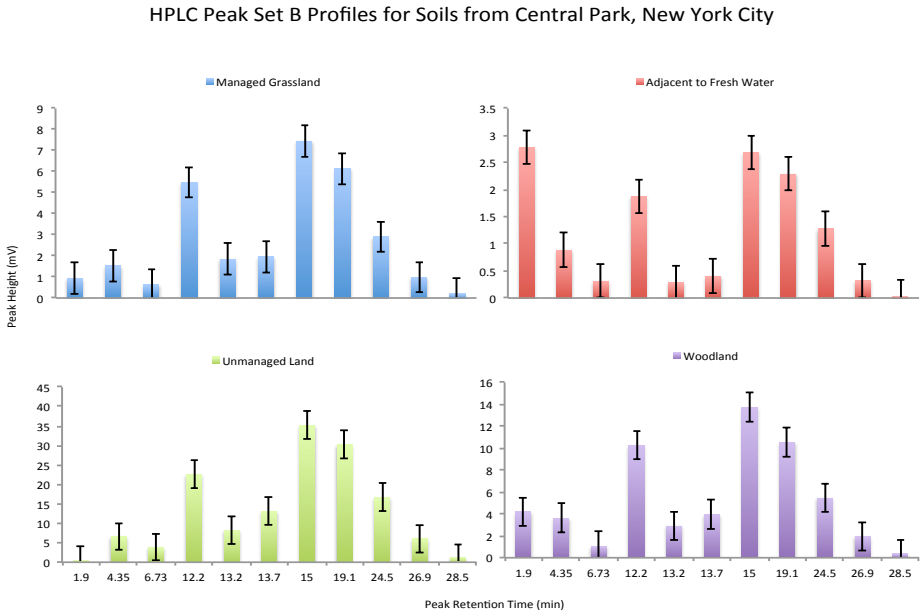
This section outlines the results of the visual comparison of the Peak Set B profiles obtained at Central Park, New York City and the results of the CDFA using the peaks in set B as variables for the Winter time point only

As has been outlined in Chapter 5, the profiles varied across the four locations in Central Park, New York City for peak set B (Figure 6.59). Soil profiles for unmanaged land could be discriminated from the other three locations by the absence of the peak at 1.9min while the size order of the peaks at 1.9, 4.35 and 6.73min allowed samples from managed grassland to be visually separated from those for woodland.

The ratio of the peak at 1.9min compared to the peak at 12.2 min allowed the profiles of soil adjacent to fresh water from the other locations, the peak at 12.2min was 1.87mV and was 33% smaller than the 2.78mV peak at 1.9min for soils adjacent to fresh water, while for managed grassland the peak at 12.2min was far larger than the peak at 1.9min with peak heighs of 5.49mV and 0.94mV, and for woodland the peak heights were 10.3mV and 4.19mV, respectively.

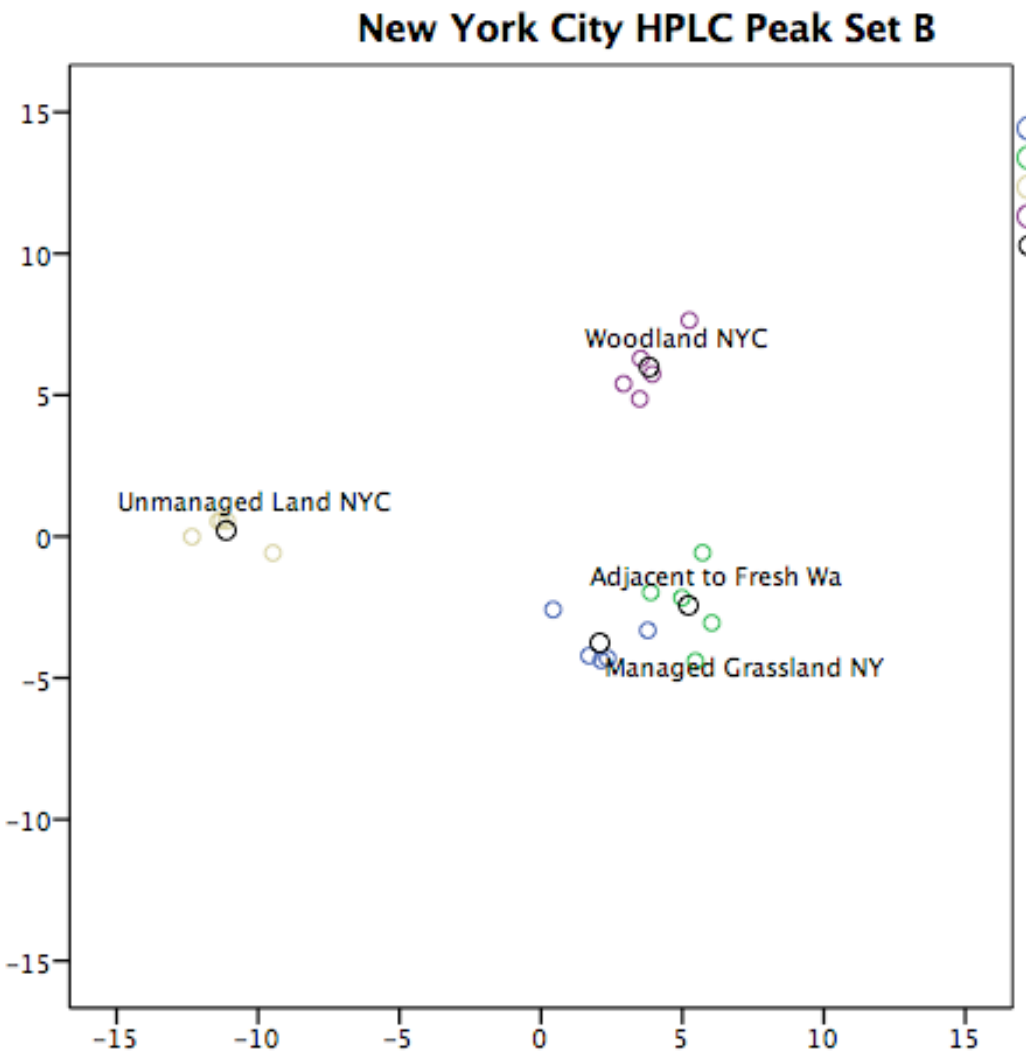
The peak sizes were far larger for unmanaged land and were smallest at the location adjacent to fresh water, while the peaks in the managed grassland samples were generally three times as large as those for the location adjacent to fresh water, and woodland samples were approximately twice the size obtained for managed grassland.

Figure 6.59: HPLC Profiles for Central Park, New York City- Peak Set B



The CDFA classified the samples from New York City with 100.0% accuracy using peak set B (Figure 6.60, Table 6.3), and the first three functions explained 73.3%, 24.1%, and 2.5% of the variability in the sample groupings. This discrimination was statistically significant at the 99% confidence level (p=0.000)

Figure 6.60 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set B- New York City



6.3.5 Summary of Seasonal Changes

The CDFA results for the analyses at each site for each of the different seasons are summarised below. Full details of the statistical analyses are provided in table 6.2 for peak set A and in table 6.3 for peak set B

For Peak Set A (Table 6.2) the CDFA produced functions that were statistically significant at the 99% confidence interval at all sites across all time points, except for Edinburgh at the Spring time point, where the discrimination was significant at the 95% confidence interval. The grouping accuracy obtained in the CDFA using peak set A was 100% for all sites at all time points except for the Autumn 2013 time-point in London, where accuracy was 89.5% and for the Aberdeen site in Winter, where samples were grouped with 94.7% accuracy.

The CDFA results for Peak Set B (Table 6.3) gave rise to discriminations that were statistically significant at the 99% confidence interval at all time-points and sites except Edinburgh in Spring, where the discrimination was statistically significant at the 95% confidence interval, and at the London site in Autumn 2013 and Autumn 2014, and the Aberdeen site in Autumn 2014, where the discriminations were not statistically significant.

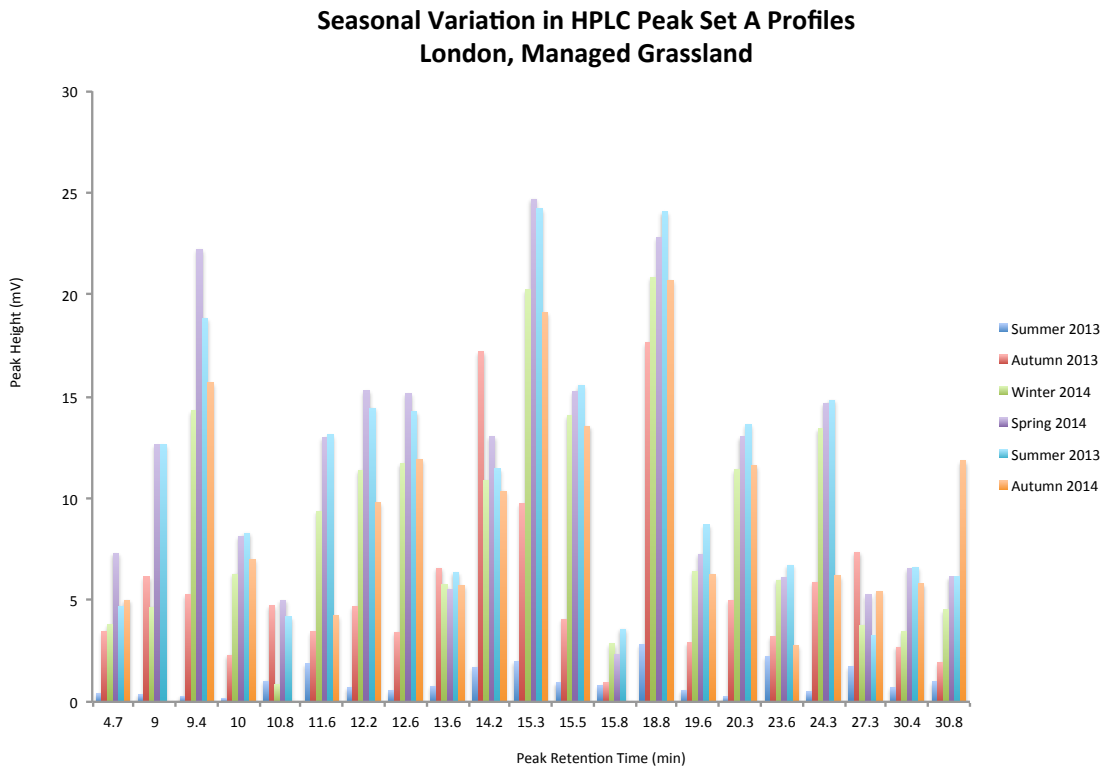
Using HPLC Peak Set A, all but three samples were correctly classified with misclassifications at three sample locations, over two time points, while 9 sample groups were misclassified when peak set B was used. The misclassified samples are summarised in Table 6.4

Table 6.4 Summary of Misclassified Samples

Time Point	Sample Site	Original Location	Peak Set	Number of Samples Misclassified as			
				Managed Grassland	Adjacent to Fresh Water	Unmanaged Land	Woodland
Autumn 2013	London	Managed Grassland	A	N/A	0	1	0
Autumn 2013	London	Unmanaged Land	A	1	0	N/A	0
Winter	Aberdeen	Unmanaged Land	A	0	0	N/A	1
Autumn 2013	London	Managed Grassland	B	0	0	1	N/A
Autumn 2013	London	Adjacent to Fresh Water	B	0	N/A	2	0
Autumn 2013	London	Unmanaged Grassland	B	0	1	N/A	0
Winter	London	Managed Grassland	B	N/A	0	1	0
Winter	London	Adjacent to Fresh Water	B	1	N/A	0	0
Autumn 2014	London	Managed Grassland	B	N/A	1	0	0
Summer	Edinburgh	Woodland	B	0	0	1	N/A
Autumn	Aberdeen	Managed Grassland	B	N/A	1	0	0

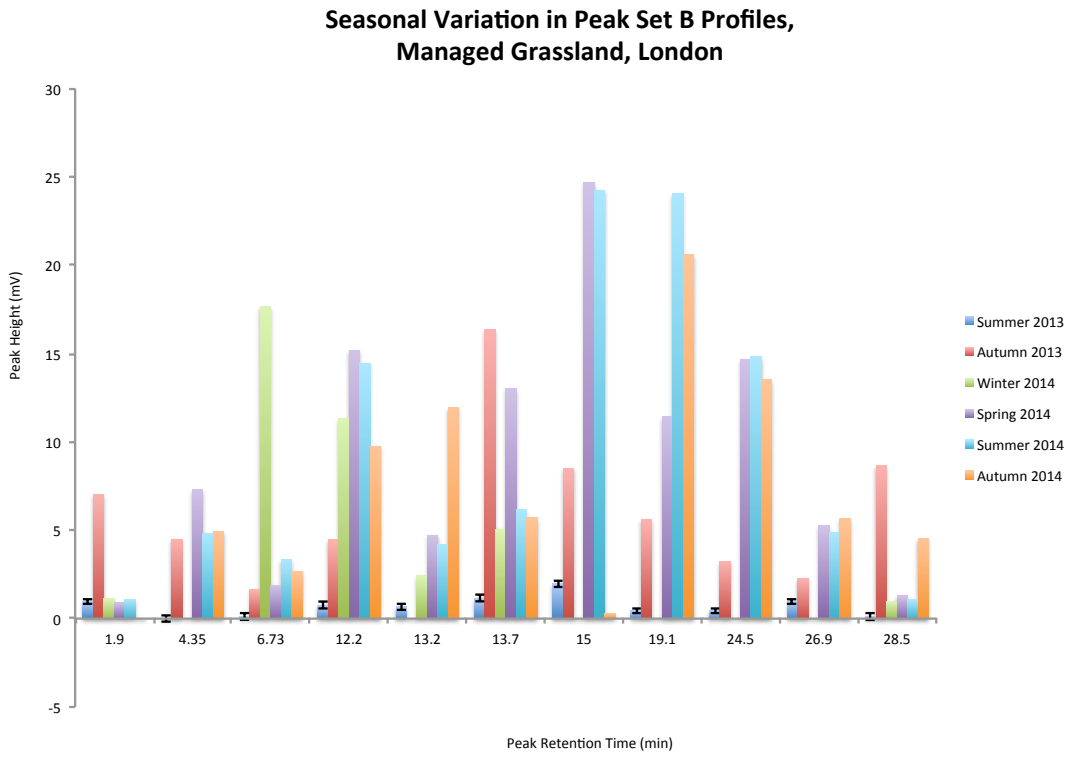
There were no seasonal markers identified through visual analysis for peak set A at any of the sites, or locations within the site, since the profiles differed at each location at each successive time point, and this variability is illustrated in figure 6.61 which presents the complete set of profiles obtained for managed grassland at the Brockwell Park site as an example, however the variability between seasons and lack of location or seasonal marker peaks was consistent across all sites and locations.

Figure 6.61 Summary of Seasonal Changes for HPLC Peak Set A, at the Managed Grassland location in Brockwell Park



It was not possible to identify seasonal markers by visual analysis for peak set B at any of the sites, or locations within the site, the profiles obtained for each location changed from one time point to the next and there were no consistent trends in the data, an example of this point is provided in figure 6.62 using the data from the managed grassland location at Brockwell Park.

Figure 6.62 Summary of Seasonal Changes for HPLC Peak Set B, at the Managed Grassland location in Brockwell Park



6.3.6 Robustness to Delays in Sampling

In forensic practice, there can be no control over the delay between the time a crime was committed, and the time the crime was detected and the relevant samples being collected and analysed. As can be seen from the profiles observed over the different time points in this study, the amount of each compound present in the soil can vary from one time point to the next, which affects the relative heights of each peak and the appearance of the profile. Therefore, it is important to ensure that the accuracy of the discrimination is not adversely effected by the changes in the profiles of the soils that occur between the time the crime is committed, when soils are transferred to an item of evidence, and the time of collection of the reference samples by investigators.

In order to determine whether there remained a sufficient level of consistency amongst the profiles of samples collected at the same location throughout year to reliably discriminate the different locations, and thereby verify the robustness of the technique to delays in sampling (the ability of the method to provide accurate discrimination despite differences in the collection time of the samples of interest) all the samples from each of the time points were

grouped together for each location and were analysed by CDFA at each of the three sites that were sampled at more than one time point, and the results are provided in Table 6.5 below.

Table 6.5 Canonical Discriminant Function Results for Robustness to differences in Sampling Time (*These results omit the samples analysed in the UCL School of Pharmacy, eliminating variability arising from the use of different instrumentation)

HPLC Profiles	Classification Accuracy %	Wilks' Lambda Significance test of functions			% Variance Function 1	% Variance Function 2	% Variance Function 3
Peak Set A *		1-3	2-3	3			
London 2013-2014	68.9 (* 91.3)	0.000 (0.000)	0.000 (0.000)	0.090 (0.030)	53.5 (57.7)	32.8 (26.9)	13.7 (15.4)
Edinburgh 2014	98.4	0.000	0.000	0.002	61.9	22.5	15.6
Aberdeen 2014	94.9	0.000	0.000	0.000	53.7	30.4	15.9
Peak Set B							
London 2013-2014	60.0 (68.8)	0.000	0.360 (0.015)	0.334 (0.217)	81.5 (76.7)	9.8 (16.1)	8.7 (7.3)
Edinburgh 2014	70.8	0.000	0.022	0.385	78.1	16.4	5.5
Aberdeen 2014	64.5	0.000	0.384	0.687	91.4	6.9	1.7

6.3.6.1 Robustness to seasonal change for Peak Set A

This section will discuss results of the CDFA for the peak set A data set, using the samples from the same site and location at different times of year, in order to assess the ability of this peak set to provide reliable discrimination when the samples have been collected at different times of the year.

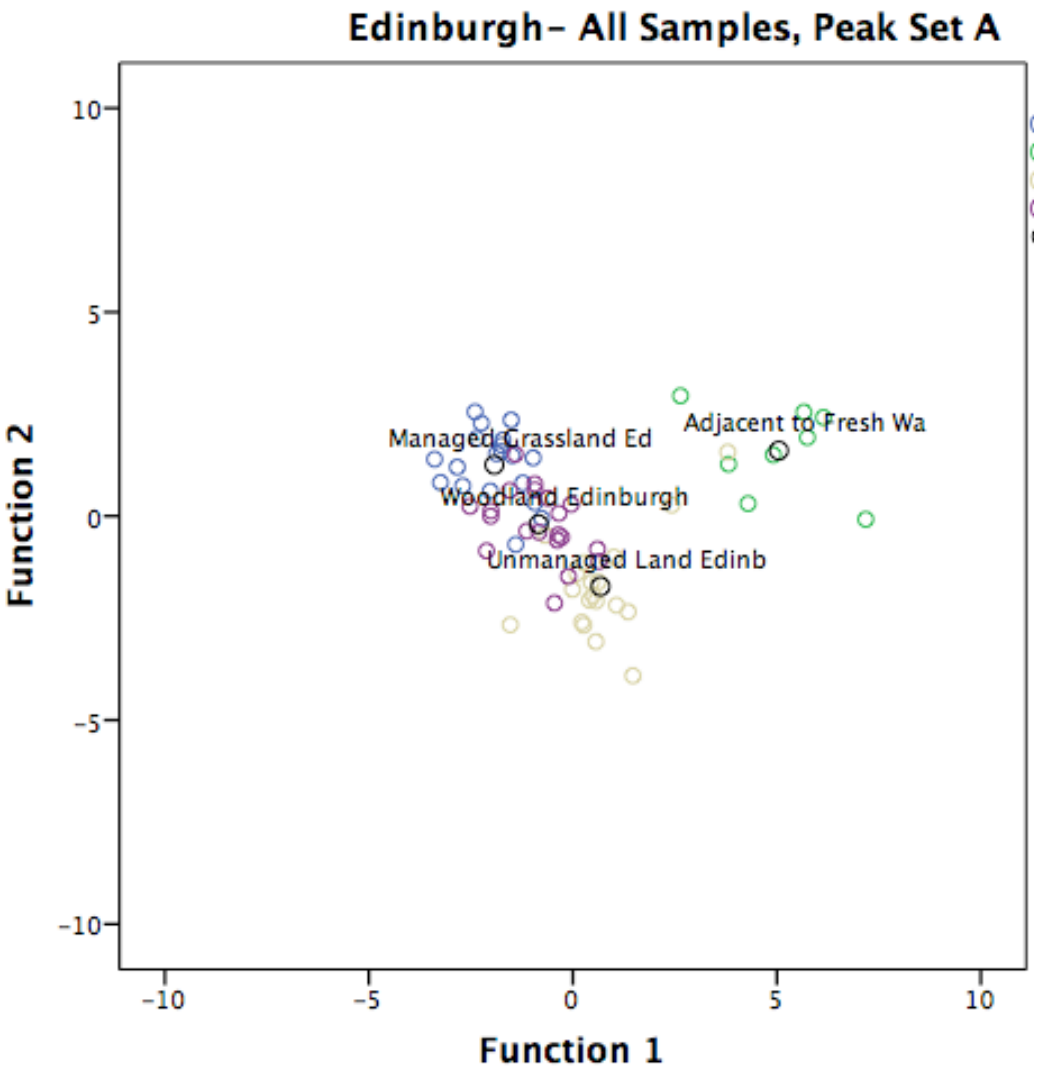
The accuracy of the classification was reduced when the sample location groups were collated from samples collected at different times of year, compared to the discrimination of the soils from locations that were sampled at the same time of year. The accuracy with which the combined samples were assigned to their location using the functions produced in the CDFA ranged from 68.9% to 98.4%, and there were sample misclassifications observed at 11 out of the 16 locations for peak set A, which are summarised in Table 6.6.

Table 6.6 Summary of Misclassified Samples for Peak Set A

Time Point	Sample Site	Original Location	Peak Set	Number of Samples Misclassified as			
				Managed Grassland	Adjacent to Fresh Water	Unmanaged Land	Woodland
All	London	Managed Grassland	A	N/A	1	8	0
All	London	Adjacent to Fresh Water	A	5	N/A	6	2
All	London	Unmanaged Land	A	3	2	N/A	0
All	London	Woodland	A	5	4	1	N/A
2014 Only	London	Managed Grassland	A	N/A	1	1	0
2014 Only	London	Adjacent to Fresh Water	A	3	N/A	2	0
All	Edinburgh	Managed Grassland	A	N/A	0	0	2
All	Edinburgh	Unmanaged Land	A	0	1	N/A	1
All	Edinburgh	Woodland	A	0	0	1	N/A
All	Aberdeen	Managed Grassland	A	N/A	0	0	2
All	Aberdeen	Adjacent to Fresh Water	A	0	1	N/A	1

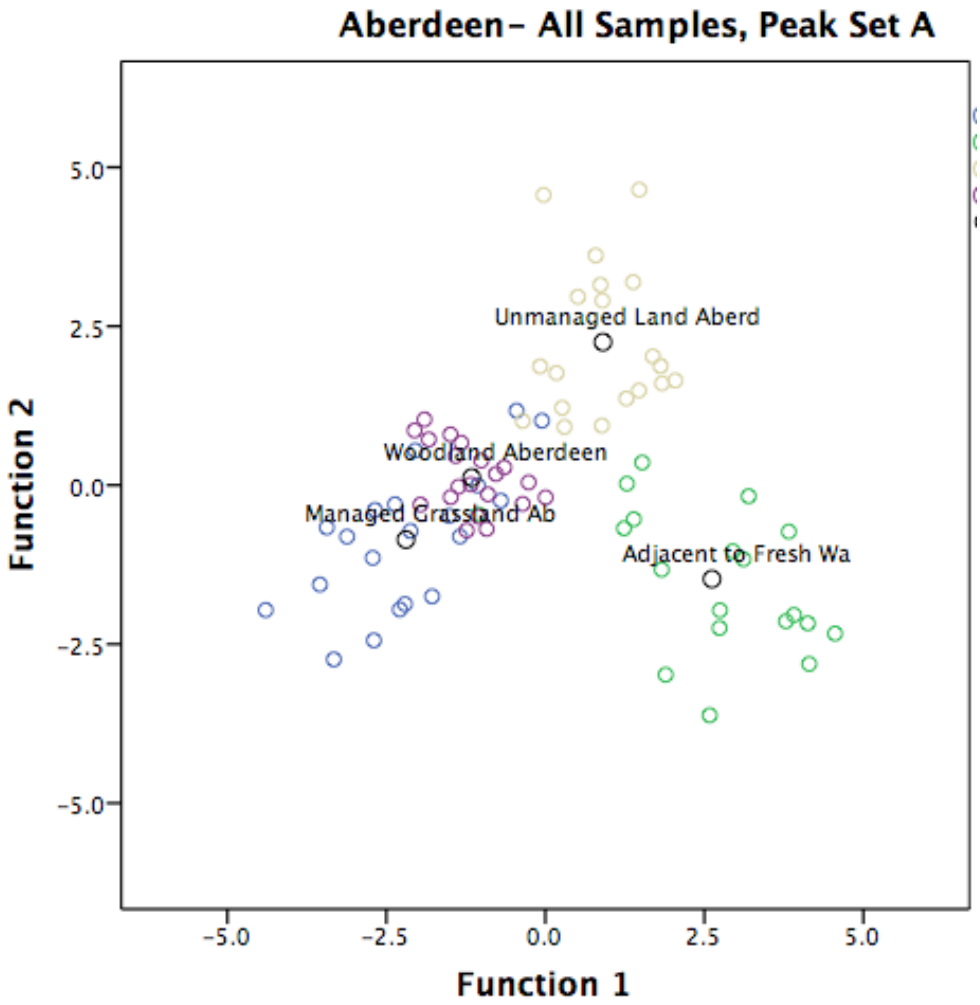
Figures 6.63 to 6.66 show the scatter plots displaying the scores for each individual sample, and the relative positions of the group centroids for each group, for the first two canonical functions obtained in the CDFA using the Edinburgh, Aberdeen, London, and London (2014 only) data sets for peak set A.

Figure 6.63 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set A- Edinburgh, All time points



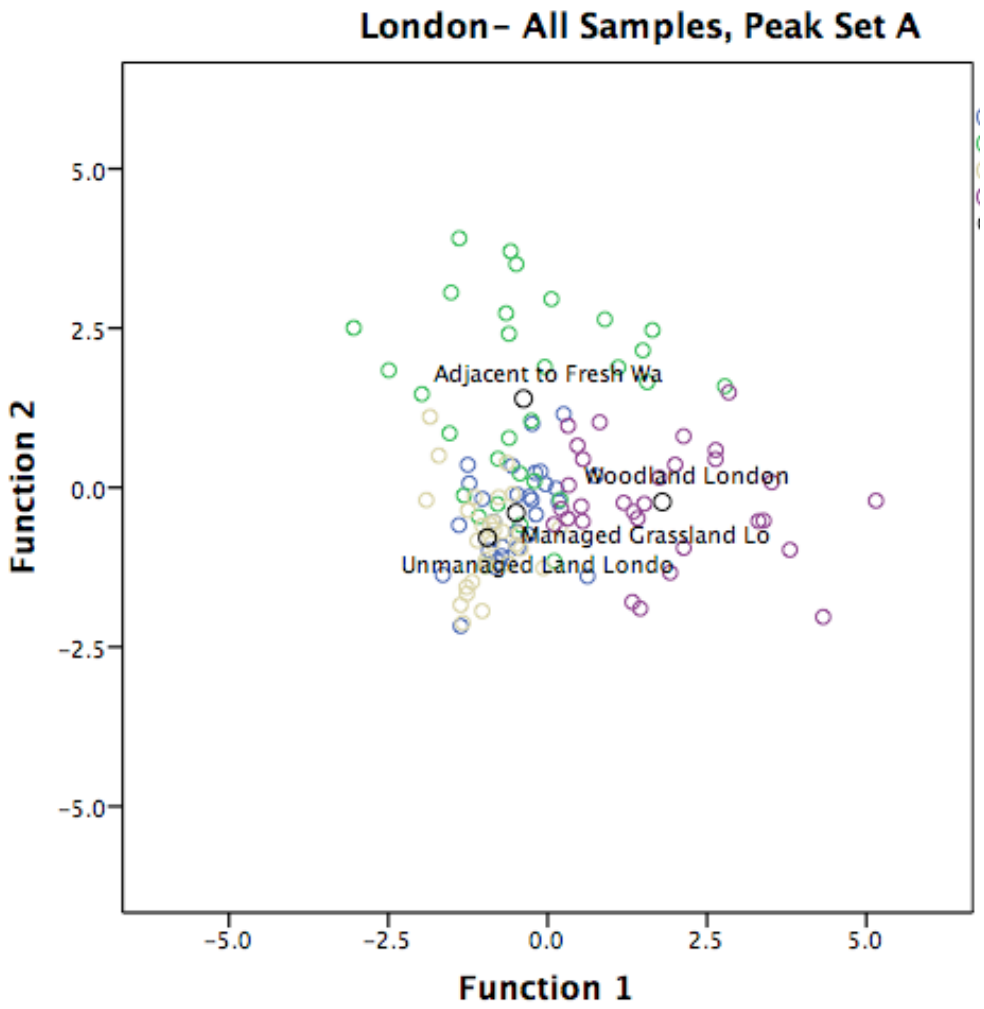
The collection of samples at different times of year gave discrimination that was 98.4% accurate for the Edinburgh site for the peak set A data (Figure 6.63, Table 6.5), and the discrimination was significant at the 99% confidence interval (p=0.000), with the first three canonical discriminant functions accounting for 61.9%, 22.5% and 15.6% of the variation between the samples from all time points.

Figure 6.64 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set A- Aberdeen, All time points



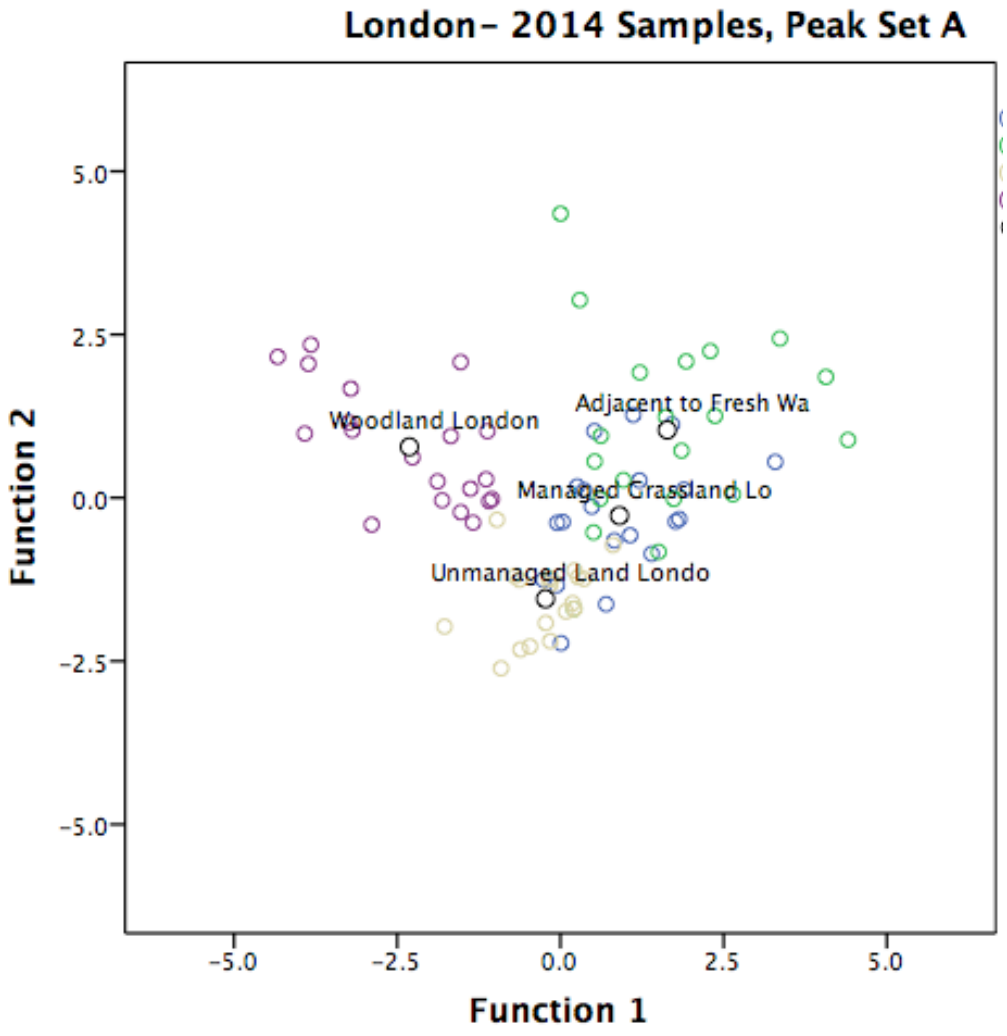
For the Aberdeen site, collecting samples at different times of year produced a discrimination that was 94.9% accurate for the peak set A data (Figure 6.64, Table 6.5). The first three canonical functions accounted for 53.7%, 30.4%, and 15.9% of the variability in the samples and the discrimination was significant at the 99% confidence interval (p=0.000)

Figure 6.65 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set A- London, All time points



The accuracy in discrimination obtained for the London data collected across all six time points was far poorer than for the Edinburgh and Aberdeen sites when peak set A was used in the CDFA, at only 68.9% (Figure 6.65, Table 6.5). The discrimination was, however, statistically significant at the 99% confidence interval (p=0.000) and the first three canonical functions explained 53.5%, 32.8% and 13.7%.

Figure 6.66 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set A- London, 2014 time points only



The poor performance of the discrimination in the London peak set A data could be explained by the fact that the analyses were performed in different laboratories and with different equipment introducing additional precision errors to the data, therefore the samples from the initial two time points, that were analysed in the UCL School of Pharmacy laboratory, were removed from the data set in order that like-for-like sample comparisons could be made during the CDFA on the data generated from the 2014 time points only, which were all tested at the James Hutton Institute laboratory (Figure 6.66, Table 6.5). The accuracy of discrimination was improved by 22.4% when the CDFA was limited to the samples that were analysed on the same instrument, and 91.3% of these samples were correctly classified with

this data set. This discrimination was also statistically significant at the 99% confidence interval (p=0.000) and the first three canonical functions explained 57.7%, 26.9% and 15.4% of the variability in the samples.

6.3.6.2 Robustness to seasonal change for Peak Set B

This section outlines the results obtained in the CDFA when the peak set B data were pooled from across multiple time points for each location and site, and the assesses the ability of peak set B to discriminate between locations in cases where reference samples have been collected at a number of different points in the year.

The accuracy of the classification was more adversely affected by the seasonal changes in the profiles for peak set B, with lower accuracy rates at all sites and with more sample locations misclassified, in comparison to peak set A. For peak set B, the accuracy with which the samples were assigned to their groups ranged from 60.0% to 70.8% and there were sample misclassifications observed for 15 out of the 16 sample groups in this data set, which are summarised in Table 6.7, below.

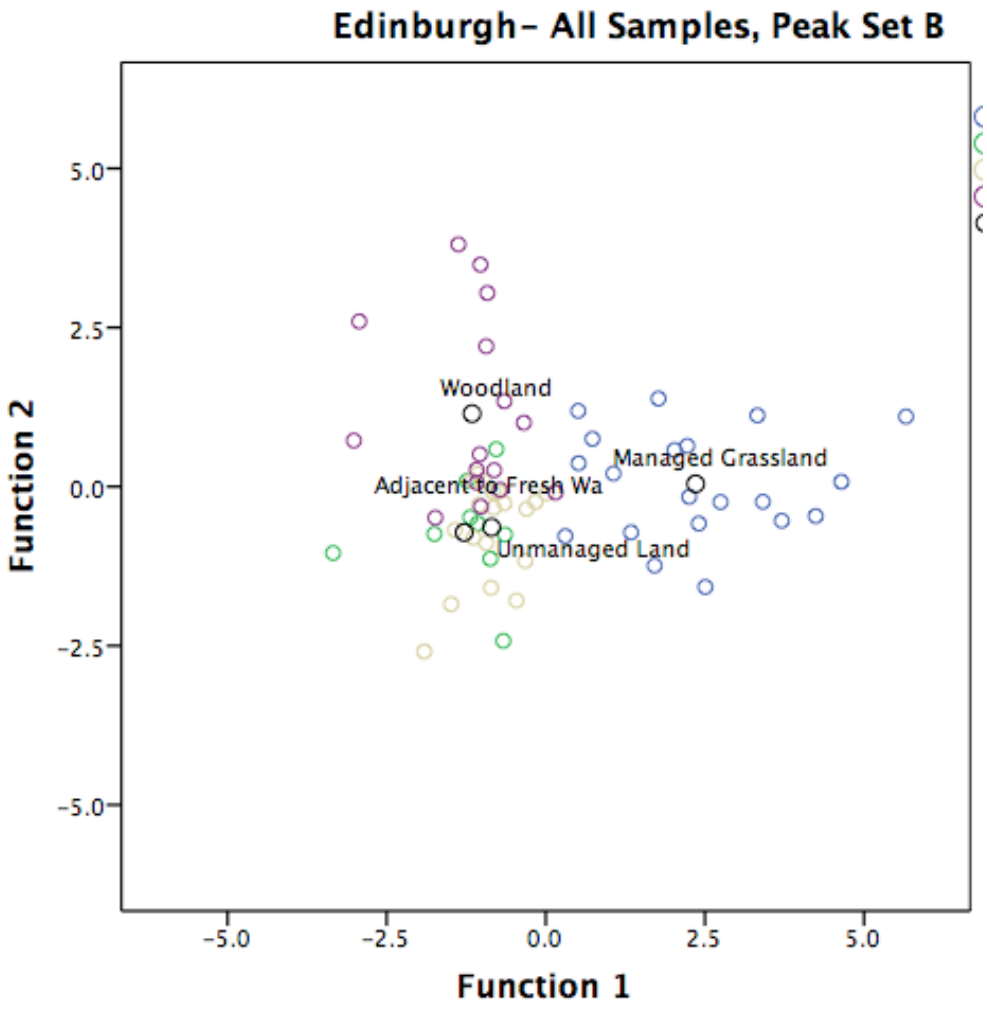
Table 6.7 Summary of Misclassified Samples Across All Time points- Peak Set B

Time Point	Sample Site	Original Location	Peak Set	Number of Samples Misclassified as			
				Managed Grassland	Adjacent to Fresh Water	Unmanaged Land	Woodland
All	London	Managed Grassland	B	N/A	8	10	2
All	London	Adjacent to Fresh Water	B	4	N/A	8	2
All	London	Unmanaged Land	B	1	0	N/A	4
All	London	Woodland	B	7	0	2	N/A
2014 Only	London	Managed Grassland	B	N/A	5	5	0
2014 Only	London	Adjacent to Fresh Water	B	2	N/A	5	1
2014 Only	London	Unmanaged Land	B	1	0	N/A	4
2014 Only	London	Woodland	B	0	1	1	N/A
All	Edinburgh	Managed Grassland	B	N/A	0	1	2
All	Edinburgh	Adjacent to Fresh Water	B	0	N/A	3	1
All	Edinburgh	Unmanaged	B	0	6	N/A	0

		Land					
All	Edinburgh	Woodland	B	0	3	3	N/A
All	Aberdeen	Managed Grassland	B	N/A	0	3	6
All	Aberdeen	Unmanaged Land	B	7	0	N/A	7
All	Aberdeen	Woodland	B	2	0	2	N/A

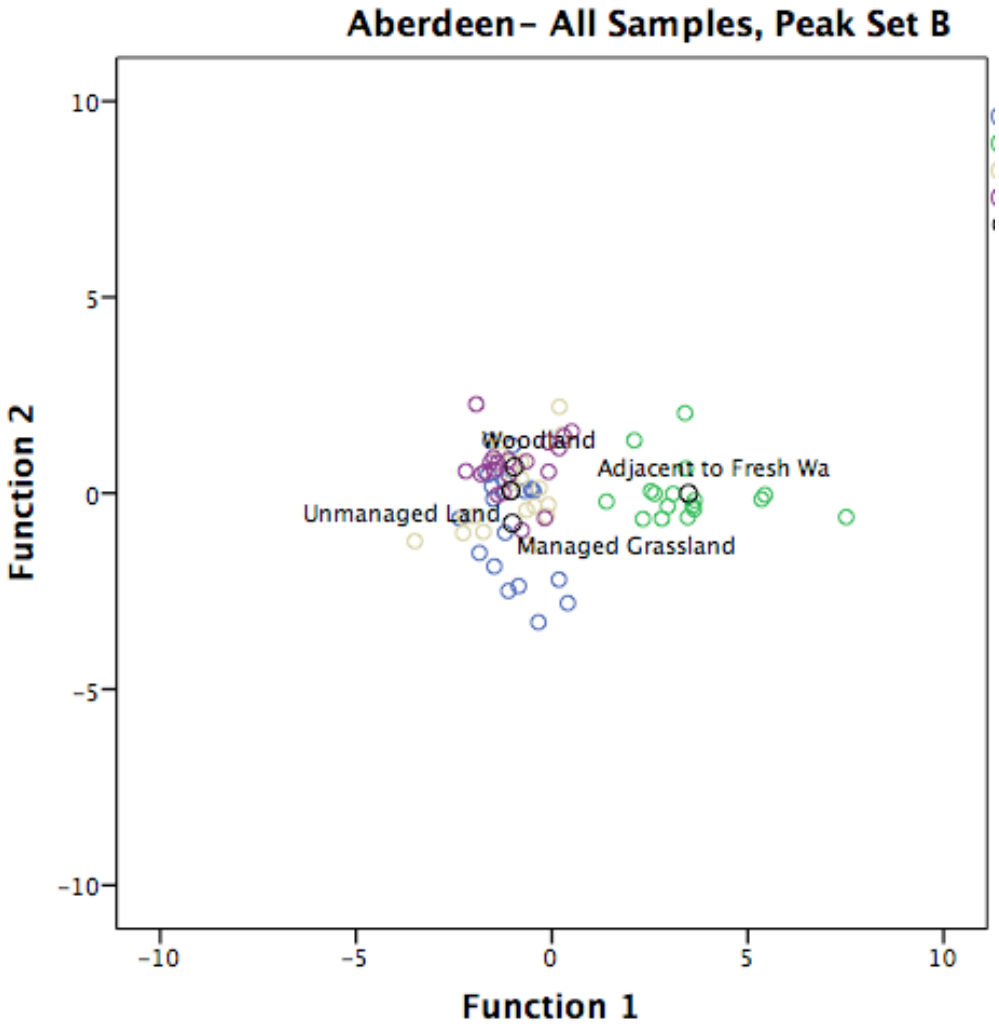
Figures 6.67 to 6.60 present the scatter plots obtained in the CDFA using HPLC peak set B for the Edinburgh, Aberdeen, London (all time points), and London (2014 only) data sets, which display the scores for each individual sample, and the relative positions of the group centroids for each group, for the first two canonical functions

Figure 6.67 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set B- Edinburgh, All time points



The CDFA on peak set B gave the highest accuracy rates at the Edinburgh site (Figure 6.67, Table 6.5), with 70.8% of the samples correctly assigned to location groups based on the first three canonical discriminant functions, which explained 78.1%, 16.4%, and 5.5% of the variability in the resultant sample groupings. This discrimination was also statistically significant at the 99% confidence interval (p=0.000).

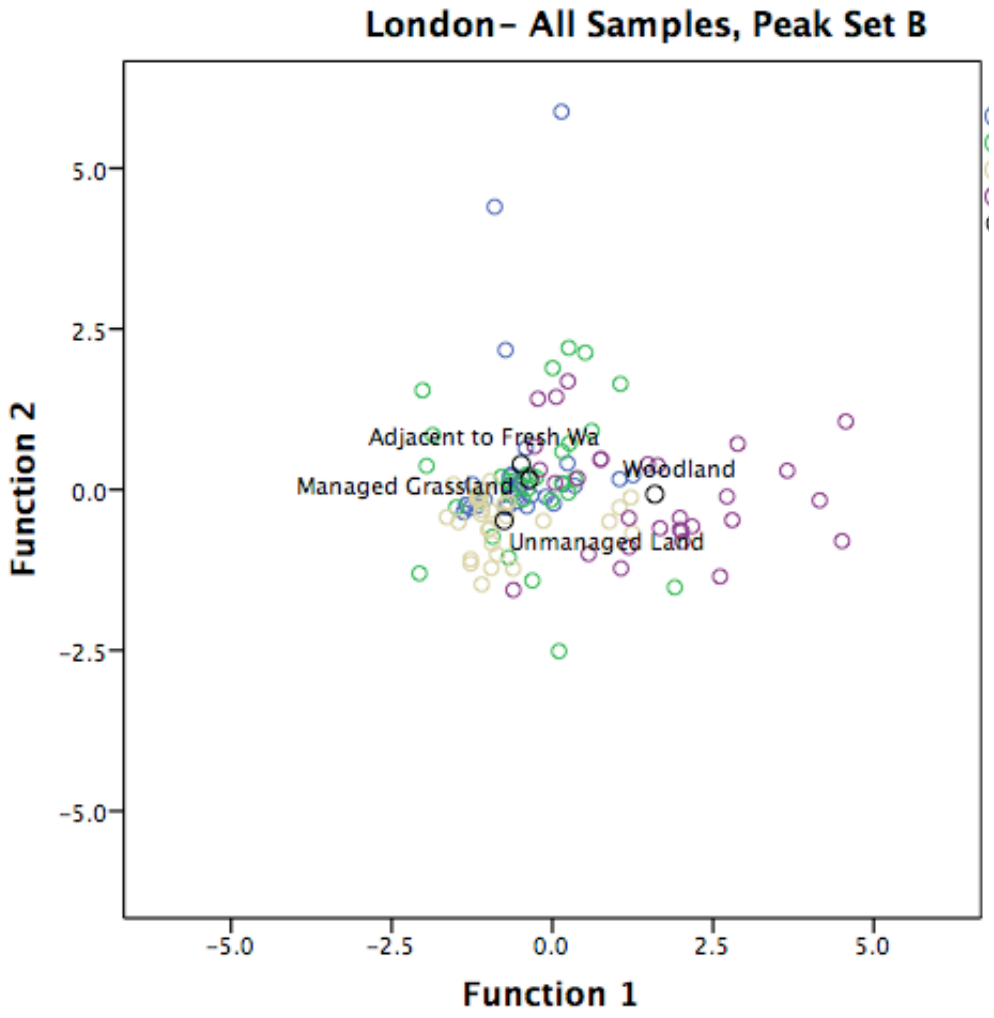
Figure 6.68 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set B- Aberdeen, All time points.



Using peak set B, the accuracy of the discrimination at the Aberdeen site (Figure 6.68, Table 6.5) was also adversely affected when the samples used in the CDFA were collected at different times of year. Just 64.5% of the samples collected were assigned to the correct group on the basis of the first three canonical discriminant functions, which were responsible for

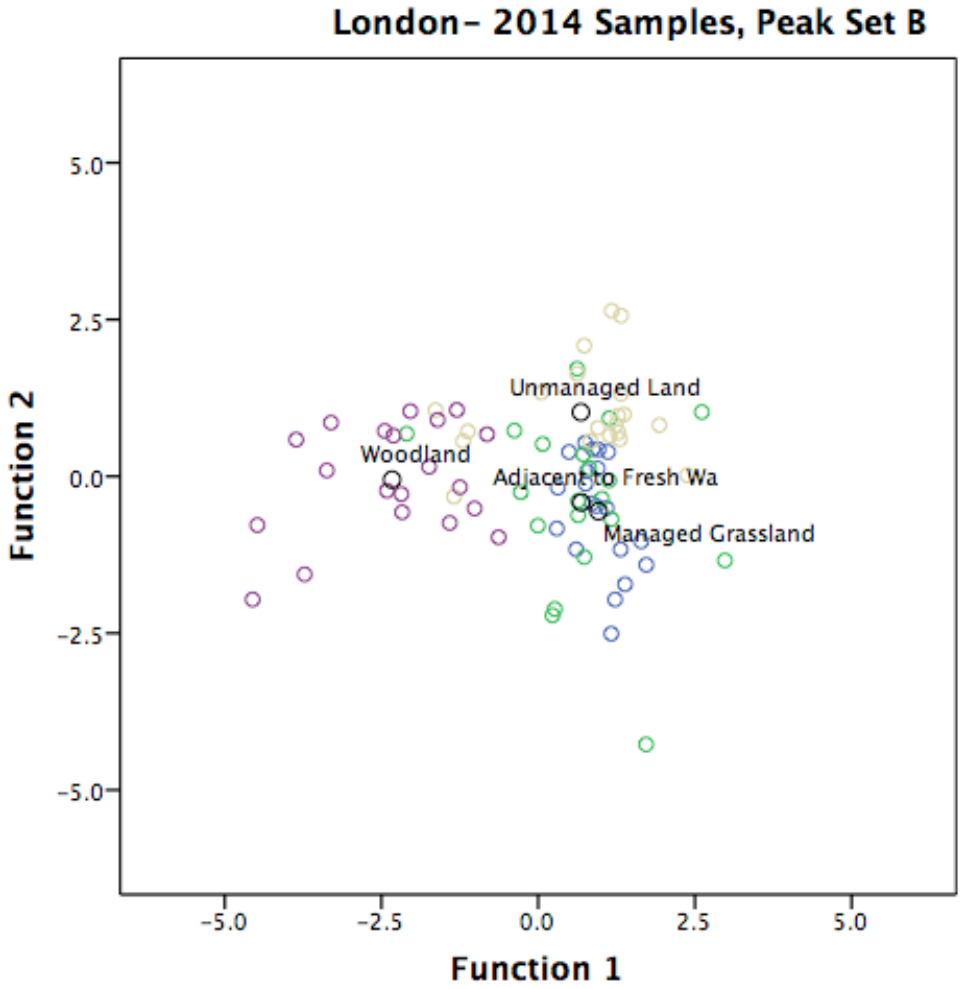
91.4%, 6.9% and 1.7% of the variation between samples in the resulting discrimination, however the discrimination was still significant at the 99% confidence interval (p=0.000)

Figure 6.69 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set B London, All time points



The results of the CDFA for London were far less accurate than for the other sites when peak set B was considered (Figure 6.69, Table 6.5), as was the case for peak set A. The accuracy for this data set was just 60.0%, however the discrimination was still statistically significant at the 99% confidence interval (p=0.000) and the first three canonical functions explained 81.5%, 9.8% and 8.7% of the variation in the samples.

Figure 6.70 Plot of sample scores for Function 2 vs. sample scores for Function 1 for Peak Set B- London, 2014 time points only



As discussed in section 6.3.6.1 for peak set A, the poor performance of the CDFA for London could potentially be caused by reduction in precision within sample groups as a result of the use of a different laboratory and equipment for the 2013 samples and the 2014 samples. Upon removal of the samples analysed at the UCL School of Pharmacy, the accuracy for the London data set increased to 68.8% (Figure 6.70, Table 6.5). This discrimination was statistically significant at the 99% confidence interval (p=0.000) and the first three canonical functions explained 76.7%, 16.1% and 7.3% of the variation in the samples.

6.4 Discussion

Peak sets A and B both provided several ways, such as peak ratios or the presence or absence of a particular peak, to distinguish the four close-proximity locations at each site based on visual assessment of the profiles, and gave high levels of accuracy when used as variables in CDFA. This section discusses the impact of varying the location and time, respectively, at which samples were collected upon the ability to discriminate the different locations and the implications for forensic case work.

The accuracy rates were high across all of the sample sites, and the majority of the sample sets that were taken at a single time point could be discriminated with 100% accuracy, regardless of the geographic location of the site. The Edinburgh samples could be separated with 100% accuracy for seven out of the eight data sets collected, while at the Aberdeen site six out of the eight data sets gave 100% accuracy, while for the London site all samples were correctly classified for eight out of twelve data sets. Both of the New York data sets yielded 100% accuracy, however it was not possible to assess whether these accuracy rates would have been obtained upon sampling multiple time points. The poorest accuracy of discrimination was observed at the London site, at 80% for peak set B in Autumn 2013, however the average accuracy rates at each of these sites were much closer, and ranged from 96.2% for the London site to 100% for the New York site, while the averages for Edinburgh and Aberdeen sites were 99.3% and 98.7%, respectively. The similarity in the average accuracy rates for the four sites implies that the ability to produce highly discriminatory profiles is not dependent on the geographical location, which corroborates the findings outlined in Chapter 5 and provides further confirmation that the technique can be considered robust to changes in underlying geology and fit for use at a variety of geographical locations throughout the UK and internationally.

There were no features in the profiles for either peak set, such as peak height ratios or the presence or absence of certain peaks, that remained constant for any site at any location across all the various time points, as shown in figures 6.61 and 6.62 for the managed grassland data at Brockwell park, so it was not possible to identify any markers that were specific to a particular geographic location, for instance there was no feature that was consistently observed at the Brockwell Park site, nor were any marker peaks identified for specific land use types, therefore there is no evidence at this stage to suggest that the HPLC profile of a location is indicative of the land use. Likewise, it was not possible to discern any commonalities between samples from the same location type at the different sites, for instance managed grassland soils were not observed to share similar profiles across all four parks. It is unlikely that geoforensic evidence from a particular location type at another site would yield an HPLC

profile analogous to that of soils with the same location type in this study, which demonstrates the need for cautious interpretation of the results for these samples and careful consideration must therefore be given to the suitability of using geoforensic HPLC profiles in conjunction with large scale databases for intelligence purposes. As discussed in Chapter 5, the lack of specific markers for a geographical location or type of land use suggests that this profiling technique is better suited for evidentiary analyses for the comparison of samples with questioned and known provenance, as opposed to the identification of a particular location type or geographical region through comparison of the profile obtained from of a piece of evidence with reference samples in a database.

In cases where all samples were collected at the same time, all four location types were grouped with high levels of accuracy, 96.7% of managed grassland samples were assigned correctly, 97.7% of samples from the location adjacent to fresh water were accurately attributed to that location, 98.0% of samples from the unmanaged land were correctly classified, while 99.3% of woodland soils were attributed to the correct location type. For the CDFA performed on samples collected across all time points, however, overall accuracy rates from for each location type were lower, and more varied. For these data the poorest accuracy was achieved for the managed grassland samples, at 68.8%, while this improved to 74.4% accuracy for samples from the location adjacent to fresh water and 79.4% for the samples from unmanaged land, while 82.2% of the soils from the woodland location were able to be assigned to the correct location. Regardless of the length of the time period over which samples were collected, the discriminatory ability of the technique across the different land uses was the same, woodland samples were the most successfully grouped location type, followed by the unmanaged land, then the location adjacent to water, with the fewest samples correctly grouped at the managed grassland location. While this suggests that the technique performs best when grouping samples from woodland locations, the ability to discriminate between different location types remained high for all analyses, particularly where the sample collection time frame is constrained, therefore the technique is not considered to be limited to use at a particular location type.

The accuracy rates were high across all of the time points examined in this study, however the accuracy rates were slightly higher for peak set A for all seasons. The poorest accuracy rate was observed at the Autumn 2013 time point using peak set B, at just 80.0%, and the poorest average accuracy rate was obtained using both peak sets from samples collected in Autumn, which was 94.9%. In comparison, the average results for all data sets at the Winter, Spring and Summer time points were slightly better, since these data sets achieved 98.1%, 100% and 99.3% accuracy in grouping the samples, respectively. In Winter, two of the six data sets

failed to give 100% accuracy, for the Aberdeen site using peak set A the accuracy was 94.7% giving an overall accuracy in grouping samples collected in Winter of 98.2% for peak set A, while for the London site using peak set B the accuracy was only 90%, giving an overall accuracy of 96.7% for peak set B in Winter. In Spring, all samples were correctly classified with both peak sets, while in Summer 100% of samples were accurately grouped for seven of the eight data sets, with the Edinburgh site achieving 94.7% accuracy using peak set B. Four of the eight data sets prepared from samples collected in Autumn were classified completely correctly, however the London 2013 data for peak set A gave only 89.5% accuracy and for peak set B the accuracy rates were just 80.0%, 95.0% and 95.0% for the London 2013, London 2014 and Aberdeen samples, respectively.

The data show that on average, the overall performance of the technique was largely unaffected by the season in which samples were collected, therefore the ability of the technique to discriminate samples from different locations is not strictly dependent on the season in which samples are collected and technique has been demonstrated as suitable for use throughout the year. There were a number of ways to visually discriminate the profiles of the four locations at each site for each of the different time point, such as comparison of the retention times of the largest peaks, however examination of the profiles obtained at each of the time points failed to reveal any seasonal markers, or common features that could be used to identify samples collected at specific times of year. Neither were there any seasonal trends observable in the profiles, for instance the growth and decline of a particular peak throughout the year. This HPLC technique has therefore been demonstrated to be robust to changes in sampling season, for cases where all samples are collected at the same time.

The profiles obtained for any particular location within a site varied were shown to be highly variable from one point in the year to the other, which could be useful in case scenarios where it is necessary to compare samples from the same specific location that are suspected to have been transferred to an item of evidence at a different time of year from the collection of reference samples. This variability did however impact negatively on the ability to make accurate discriminations between sample locations when the samples had been collected at different times of year, although these effects varied according to sample site and the data set used. Peak set A performed far better than peak set B, with accuracies in the ranges of 68.9-98.4% and 60.0-70.8%, respectively for the two sets, which improved to 91.3-98.4% and 68.8-70.8% when the data generated in a different laboratory were eliminated, which suggests that peak set A would be a better choice in cases where there is doubt as to when the crime has been committed or there is suspected to have been a change in season prior to collecting reference samples. The Edinburgh sites produced the highest accuracy of discrimination for

these data sets (98.4%), and the lowest accuracy for the London site (68.9%) however, as discussed previously, variability may have been introduced to the London data set through the necessity to use a different laboratory and HPLC instrument at the two earliest time points, and when these potentially affected results were removed, the London site performed far better (91.3%). These results suggest that the HPLC technique is robust to delays in sample collection, provided that peak set A is used for the discrimination.

The classification accuracy rate achieved for the two sets of marker peaks was very high in both cases, with both sets achieving 100% accuracy at the majority of sites and time points, and the results for the full data set were consistent with those discussed in chapter 5 for the Winter data and were slightly better for peak set A. Peak set A which achieved 100% accuracy in discriminating sample groups for 13 out of 15 the single time-point data sets, giving an overall accuracy of 99.3% for these data, while for peak set B 100% accuracy was achieved for 10 out of the 15 data sets, giving an overall accuracy rate of 96.98%. For the analyses using data collected at multiple time points, overall mean accuracy rate was once again higher for peak set A which was 88.33% than for peak set B, where the mean accuracy across the four data sets was 66.03%. There were misclassifications for all of the data sets collected at multiple time points, for both peak set A and B, however there were fewer misclassifications for peak set A than for peak set B, as 53 of the total 700 sample classifications on these data were misclassified when peak set A was used, while this figure was more than doubled when peak set B was used, resulting in 119 misclassified samples. When all data sets are considered, set A has been demonstrated to offer higher accuracy rates in all cases, and performed far better than peak set B for data sets containing samples collected at different times of year.

The error bars displayed on the profiles reflect the variability in the data for each location and, as observed for the winter data discussed in Chapter 5, the intra-location variation was similarly high for the rest of the time points, and was expected since the replicate samples had not been homogenised or pooled. These results are therefore considered to offer an accurate representation of the heterogeneity of each location of forensic interest, since any potentially important diagnostic features in the profiles have been preserved, which is crucial when comparing and excluding known and questioned samples. As discussed in Chapter 5, the variability within the very small area of soil at the individual sample points is likely to be irrelevant in many crime scenarios, therefore it should be possible to improve the precision between replicates without jeopardising accurate interpretation, by grinding the sub-samples prior analysis. This approach would not only ensure that the composition of the sub-samples were consistent, which reduces the absolute differences between the peak heights for each of the sub-samples, but also increase the amount of each compound in the final sample solutions

and therefore improve the sensitivity of the technique, which would reduce the relative magnitude of the variability in the data compared to the size of the peaks.

Another potential source of the variability observed in this study was the selection of locations at each site, as was discussed for the winter data in Chapter 5. Although there were broad qualitative similarities in the land-use for each location chosen within a site, there were no additional pre-selection criteria, such as controlling the variation and species of surrounding vegetation for each type of location, which was considered to be a more forensically relevant approach since offenders are more likely to consider situational factors such as visibility, accessibility and frequency of public usage, rather than the specific vegetation planting, when selecting a location to undertake criminal activities.

It may have been possible to maximise the inter-location variability and identify vegetation or land use markers through careful control of the planting at samples locations, for instance to ensure that specific plant species were present and abundant for the same location types at different sites, however the ecological validity of any technique developed in this manner would be limited to situations where the vegetation at the crime scenes and alibi sites was similarly constrained. In addition, the intra-location variability could have been reduced if sample locations had been chosen with limited diversity in the surface vegetation, for instance by restricting the sample sites to monoculture areas, however the degree of improvement in precision obtained using a more controlled experimental design is considered to be minimal, since in this study there were equally high levels of variability obtained for locations of homogenous vegetation. For example, the variability of the profiles generated for the managed grassland locations, where the surface vegetation (turf grass) was consistent at each sample point and each of the sites, was comparable to the results from woodland locations, where there was a variety of plants within each location and predominantly different tree species at each of the sites.

Further investigation of the complete HPLC profiles obtained in this study may be able to identify alternative sets of useful markers that are indicative of the time of year the samples were collected. In addition, future studies utilising a more frequent sampling strategy may be able to discern the rate of decay or increase of particular profile peaks in order to aid interpretation in cases where reference samples cannot be collected soon after the forensic event, or there has been a significant delay in detecting a crime or locating items of evidence.

6.5 Conclusions

The results of this study show that the ability of this technique to discriminate between locations in a small scale site is not affected by the underlying geology or geographical location of the sample site, nor by the season in which the samples are collected. The analysis was also shown to be robust to delays in the collection of reference samples, or other case scenarios where the samples are collected at multiple different times of year. There were no seasonal patterns in the data, for instance a steady increase in the size of a particular peak or peak ratio, nor were there any seasonal or location markers identified in the data, and the profiles varied at each site and location between time points, which suggests that it is not possible at this time to use the technique in "seek and find" intelligence cases. It may however be possible to identify alternative sets of markers, better suited to identifying specific time points, geographical locations, or land use, through further data analysis of the chromatographic profiles generated for this study, or through further empirical studies utilising a more frequent sampling strategy or more strictly defined location types. Nonetheless, the technique has been shown to be well suited to its intended purpose of making comparisons between samples with high rates of accuracy in performing exclusionary analyses of samples from different, close proximity locations, all year round at a range of sites in the UK and abroad.

7 Complementary Techniques

7.1 Introduction

The HPLC method developed in Chapter 3 has been shown in Chapter 4 to generate chromatograms that allow for highly accurate discrimination of close-proximity sites, and the selection of peak marker sets in Chapter 5 significantly improved the practicality of the technique by reducing the time required to clean the resulting data for exclusion by CDFA and provided two alternative analysis strategies allowing the technique to be applied in the manner most appropriate the forensic scenario. In Chapter 6, the ability of the HPLC technique was shown to be robust to various case relevant factors, such as changes in the underlying geology of the site and the time of sample collection, and could therefore potentially be applied in a range of different forensic scenarios, however the performance of this new technique has not yet been assessed in relation to the existing techniques used in forensic soil analyses described in Chapter 2. This chapter, therefore, seeks to determine the potential for this novel HPLC approach to add value to the current suite of analytical techniques available to forensic geoscientists through comparison of the results obtained using HPLC, at this close-proximity spatial scale, with two established methods that have previously used for the interpretation of soil and sediment evidence in court.

As discussed in Chapter 2, the importance of utilising independent forms of analysis in the assessment of forensic evidence has been outlined in the published literature, and it is therefore of significant value to incorporate the combined analysis of the organic fraction with the analysis of the inorganic fraction of forensic soils and sediments therefore, analysis of the inorganic fraction of the samples was also undertaken using Quartz Grain Surface Texture Analysis (QGSTA). Quartz grains are highly persistent and abundant in soil samples and analysis of quartz grain surface texture by SEM can be used to reconstruct the geological history of the quartz grains present in soils and sediments (170). The technique has previously been demonstrated to be highly informative in the investigation of the provenance of geoforensic samples (137) (106) and offers a number of advantages in forensic investigations since it requires very small quantities and simple and a non-destructive sample preparation, which aids the interpretation of data derived from samples comprised of mixtures of material from different provenances (13). QGSTA has been shown to work well in combination with other forensic analytical techniques in previous studies (106) (135) and was therefore, considered a good candidate for use in conjunction with HPLC.

It is also beneficial to be able to offer a range of analyses to quantify and profile the organic components of soil, to ensure that the organic fraction can be utilised in a wide variety of case scenarios, and as discussed in Chapter 2, there are many analytical techniques used in soil science to monitor the organic composition of soils for agricultural or environmental protection purposes, or indeed in earth sciences research. For forensic purposes however it is problematic that many of these analyses require large quantities of sample or require complex sample preparation, to the extent that they cannot give the levels of accuracy and precision, required for forensic work, nor can they be considered practical for implementation in a forensic context (14)(111).

There is, however, another chromatographic approach that has been demonstrated to offer valuable data from the analysis of the composition of the organic fraction of soil; the determination of wax markers by Gas Chromatography (GC) This technique produces complex profiles which are known to vary across small, forensically relevant spatial scales and for sites with different planting and land uses (18)(92) (108) (111) (31) (32) (30) (33) (34). The profiles of wax markers in soil have been found to reflect the composition of the compounds found in the leaves, stems and roots of the plants grown in them, and these profiles are known to remain stable over time, providing a historical record of the vegetation present at a site (111). A database of wax marker profiles has been developed for a range of forensically relevant land use and vegetation types in the UK, and the wax marker profiles of numerous plant species are now well understood, and as such, this type of analysis is potentially an excellent intelligence tool (112). In addition, since the wax marker profiles are so complex, it has been demonstrated that it is possible to profile soil from a specific location for evidentiary purposes (108) Since determination of the wax marker content by GC currently provides highly discriminatory results in forensically relevant scenarios and can be used in conjunction with existing soil databases, this was chosen as a comparator technique in this study to identify forensic scenarios where the HPLC technique could offer added benefits or where combined use of the two techniques could be used to complement one another by corroborating the results for the organic fraction, or improving the spatial precision or accuracy of the discrimination.

The aim of this study was therefore to compare the newly developed HPLC method with QGSTA and GC to evaluate the extent to which it can offer a complementary approach to enhance or augment the informative value of soils and sediments in forensic investigations. The sections of this chapter which compare the results of HPLC and QGSTA have been presented by McCulloch et. al. (33), and those sections comparing HPLC and GC have been presented by McCulloch et. al (in submission) (35).

7.2 Methodology:

The four sites described in Chapters 4-6 (Brockwell Park in London, Lochend Park in Edinburgh, Craigiebuckler Estate in Aberdeen, and Central Park in New York City) were used in this study. For the comparison with QGSTA, samples were collected from the London site and analysed as part of the proof of concept study (Chapter 4) and the methodology for these data is discussed fully by McCulloch et. al. (33), while the HPLC samples used to compare the results of the wax marker determination by GC were those samples collected from each of the four sites and analysed as described by McCulloch et al (34) in the study to identify the marker peak sets (Chapter 4) and used as the Winter time point during the investigation of geographic and temporal variability (Chapter 6).

7.2.1 QGSTA Methodology

As described in McCulloch et. al.(33), approximately 5g of each sample was allowed to equilibrate to ambient conditions then washed with deionised water until no further organic matter was visible. The remaining, inorganic fraction was transferred to a watch glass and allowed to dry. The samples were viewed under a binocular microscope (Nikon 10x/23) and 50 quartz grains were removed with forceps and placed onto a scanning electron microscope (SEM) stub covered with double sided adhesive tape. The SEM stubs were then sputter coated with gold and examined by SEM (Cambridge Instruments Stereoscan 90 at 5KeV). The morphology and surface texture of each grain was assessed and the range of grain types present identified accordingly (after Bull and Morgan (137)), each grain present in each sample was assigned a category and the relative abundance of grains in each category was used to compare the samples.

7.2.2 Wax Marker Methodology

Wax marker analysis was performed according to the method detailed in Morrison et. al. (112), which was derived from Dove and Mayes (171), excluding the derivitisation steps as no mass spectrometry was required to identify these markers. All solvents were redistilled prior to use and all aliquots were transferred with glass tipped, calibrated, auto-pipettes. All glassware was ashed and rinsed in n-heptane prior to use. All samples were crushed with a mortar and pestle, weighed, then dried in an oven at 50 ^{o}C for two hours and their moisture content calculated.

To each dry sample, internal standards of docosane (C_{22}) and tetratriacontane (C_{34}) were added at a concentration of 0.0506mg/g in decane, then n-heptacosanol (1-C_{27}-ol) was added at a concentration of 0.2179 mg/g in 50:50 (v/v) n-heptane: ethanol. Samples were saponified

in 1M ethanolic KOH at 90^{O}C for 16 hours then the organic layer was extracted in 3:1 water: n-heptane. This solution was evaporated to dryness and redissolved in heptane.

The extract was loaded onto a silica column that had been loaded with n-heptane, then the hydrocarbons were eluted with an n-heptane: ethyl acetate solvent system, ketones were removed from the column in the 97:3 fraction, and the crude alcohol fraction was collected in the 80:20 fraction. The hydrocarbon and crude alcohol extracts were then evaporated to dryness.

7.2.2.1 Alkane sample preparation

The hydrocarbon fraction was redissolved in n-heptane, evaporated to dryness, then dissolved in 20µl dodecane prior to injection onto the GC using the parameters detailed in table 7.1.

7.2.2.2 Alcohol sample preparation

Each of the crude alcohol fractions were redissolved in 100 µl n-heptane and 30µl was loaded onto a capped, 1ml glass solid phase extraction (SPE) column which was packed with 60mg Chromasorb HP (80-100 mesh) packing material in n-heptane, and to which 60µl saturated urea in ethanol had been added. The columns were placed in an oven at 70 ^{O}C for 20min and allowed to dry overnight, then the sterol/stanol fraction was recovered by elution with n-heptane and the n-alcohols were subsequently eluted with water. Both fractions were then evaporated to dryness prior to derivatisation.

Acetate derivatives of the n-alcohols were prepared by heating overnight with pyridine: acetic anhydride (5:1) at 50^{o}C. The acetate derivative was then repeatedly re-dissolved in n-heptane then evaporated to dryness until no acetic acid smell was observed, then dissolved in 25µl dodecane prior to injection onto the GC using the parameters detailed in table 7.1.

Table 7.1: GC Instrument Parameters

Sample type:	n-alkane	n-alcohol
Column	SGE BP1 0.5um, 30m x 0.52mm id.	ZB 5HT Inferno 0.25um, 30m x 0.25mm id
Temperature programme	170C for 4min, 30 ^{o}C/min to 215C, 1 min hold,	170 ^{o}C for 5 minutes 30^{o}C/min to 210C 1 min hold, 5.3^{o}C/min to 320^{o}C, 7 min hold
Injector	0.7ul, direct injection (280C)	0.8ul, direct injection (275 ^{o}C)
Detector	Flame Ionisation (340 ^{o}C)	Flame Ionisation(300 ^{o}C)
Carrier Gas	Helium, 4ml/min	Helium, 3.5ml/min

7.2.3 Data Analysis

7.2.3.1 HPLC Data analysis

The marker sets presented in chapter 5 and by McCulloch et al. (34) are useful markers for exclusionary comparison of soils at this spatial scale, which give very high accuracy when grouping samples using multivariate statistical methods. The HPLC profiles were first integrated using Agilent Chemstation software, eliminating all peaks that were below the limit of quantification, then the data for each of the peaks in Table 7.2 were adjusted for variations in sample quantity and CDFA was performed on the resulting data in SPSS.

Table 7.2: Retention Times of HPLC Markers (McCulloch et al.) (34)

Marker Set	Peak Retention Times (min)
A	4.4, 9.0, 9.4, 10.0, 10.8, 11.6, 12.2, 12.6, 13.6, 14.2, 15.0, 15.5, 15.8, 18.8, 19.6, 20.3, 23.6, 24.3, 37.3, 30.4, 30.8
B	1.9, 4.4, 6.7, 12.2, 13.2, 13.7, 15.0, 19.1, 24.5, 26.9, 28.5

7.2.3.2 GC Data Analysis

The GC data were analysed using chromquest software, then the absolute concentration of each *n*-alkane and *n*-alcohol was calculated relative to internal and external standards. Normalised concentrations were calculated relative to the total concentration of *n*-alkanes or *n*-alcohols in the sample, then the mean values and standard errors for each *n*-alkane and *n*-alcohol were calculated and plotted using Microsoft Excel. Previous work has shown that the odd chain *n*-alkanes and even chain *n*-alcohols are typically more informative and discriminatory, therefore the profiles of these markers were also plotted for comparison (109)(112). The data were then analysed by CDFA using SPSS.

7.3 Results and Discussion

The results of the comparison of the complementary use of HPLC profiling and QGSTA are presented in section 7.3.1 while section 7.3.2 discusses the comparison of HPLC profiling with wax marker profiling by GC

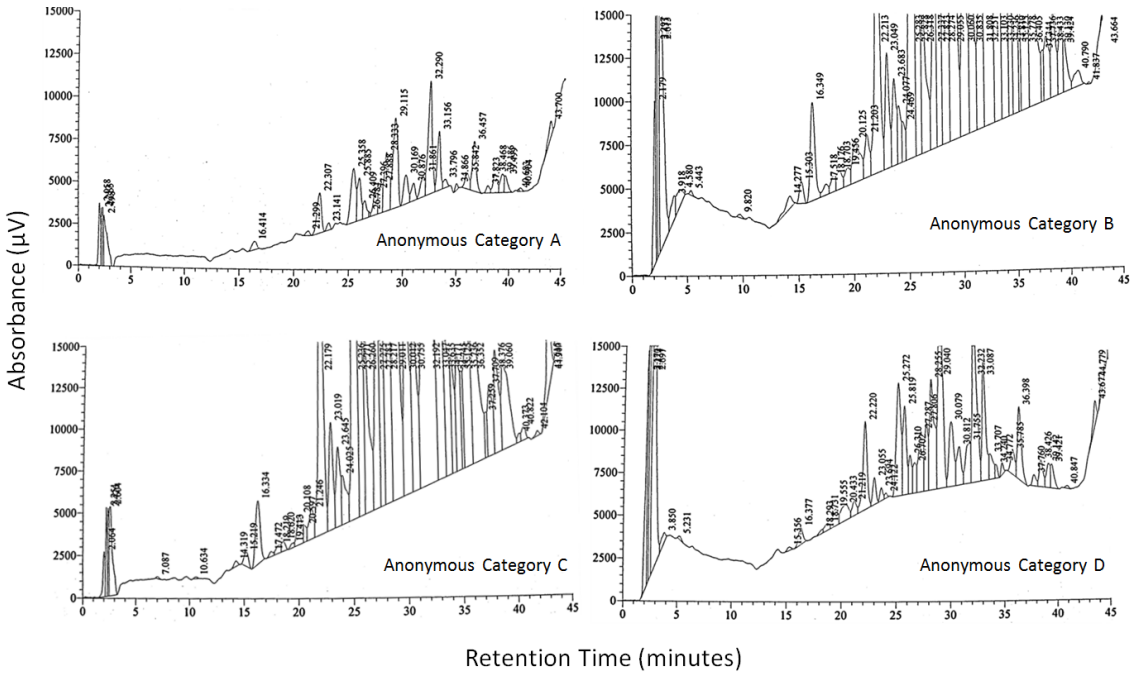
7.3.1 Comparison with QGSTA

This work was performed as part of the feasibility study outlined in Chapter 3, however the results are presented in summary form here, in order to compare the HPLC results with the QGSTA results.

7.3.1.1 HPLC

As discussed in Chapters 4 and 5, and presented by McCulloch et. al (33) distinctive chromatographic profiles were observed for the four groups of samples (Figure 7.1).

Figure 7.1: Visual comparison of chromatograms



The characteristic features detailed in table 7.3 allowed samples to be grouped with 100% accuracy by visual comparison. The process of visual comparison was relatively time consuming, at two hours for 20 samples, and naturally subjective, preventing statistical assessment of the significance of the degree of variability in the chromatography between groups and within groups.

Table 7.3: Classification of anonymised samples by visual comparison

Anonymous Category	Anonymised Samples Present	Sample Points (respectively)	Major Peaks Present		
			3-5min	17-18min	20-21min
A	5, 8, 14, 18, 19	1E, 1B, 1D, 1C, 1A	No	No	No
B	1, 2, 4, 9, 10	4E, 4B, 4D, 4A, 4C	Yes	Yes	Yes
C	6, 7, 9, 13, 20	2D, 2A, 2C, 2B, 2E	No	Yes	Yes
D	3, 11, 15, 16, 17	3C, 3E, 3A, 3B, 3D	Yes	No	Yes

The statistical anaysis by CDFA for these data (Figure 7.2) are discussed in greater detail in Chapters 4 and 5 and by McCulloch et. al (33), however the groups were separated with 100% accuracy by the three discriminant functions produced in the analysis, which explained 89.5%, 9.3%, and 1.2% of the variance between the sample groups, respectively and were significant at the 99% confidence level.

Figure 7.2: CDFA scatter plot for HPLC data

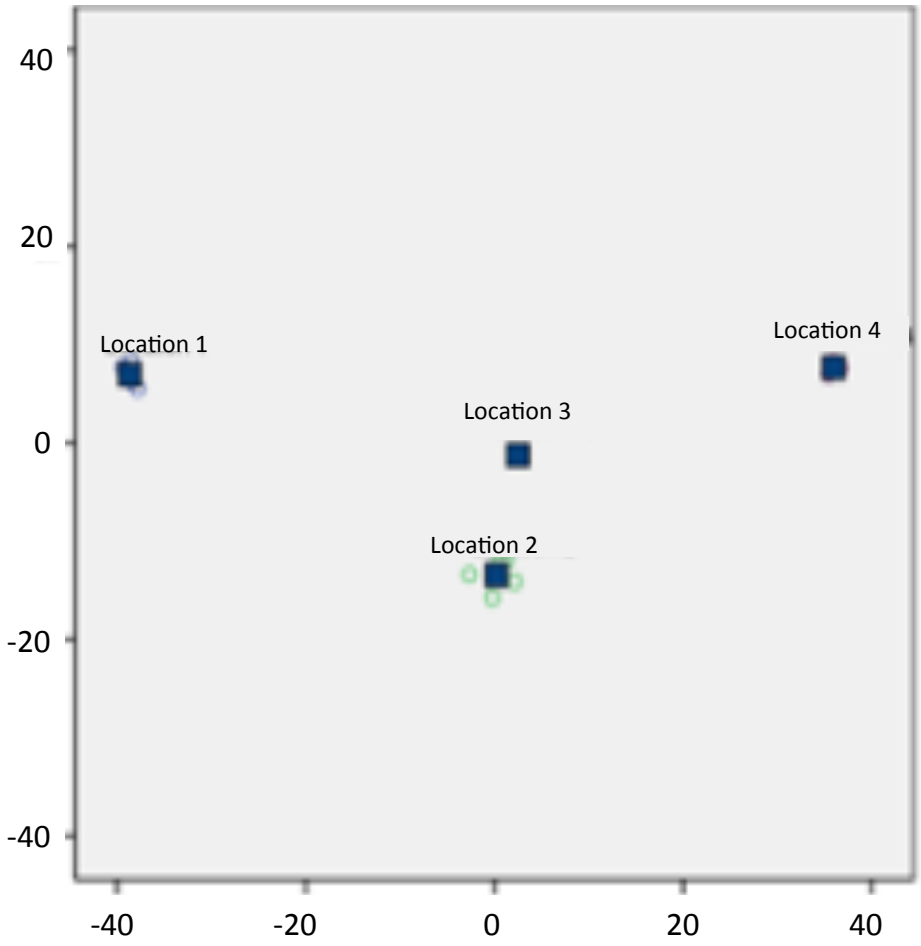


Figure 7.2: CDFA scatter plot for functions 1 and 2 showing location 1 (managed grassland) in blue, location the 2 (the location adjacent to fresh water) in green, location 3 (unmanaged land) in yellow and location 4 (woodland) in purple, group centres are indicated by blue squares.

7.3.1.2 Quartz Grain Surface Texture Analysis

Analysis of the quartz grains revealed the presence of one predominant grain type in all the samples, as has been presented by McCulloch et. al (33). This quartz grain type was a diagenetic grain with some marine and fluvial indentors with smooth etching on the surfaces. It was possible to distinguish between 3 'sub-types' of grain within these samples on the

degree of roundness (type Ia rounded grain, type Ib subrounded) and the presence of complete grain breakages (type Ic) as shown in Figure 7.3. It was also observed that many of the grains analysed were elongate as shown in Figure 7.4.

Figure 7.3: SEM images of a Type 1a rounded grain (top), a Type 1b sub-rounded grain (middle) and a Type 1c grain displaying complete grain-breakage (bottom).

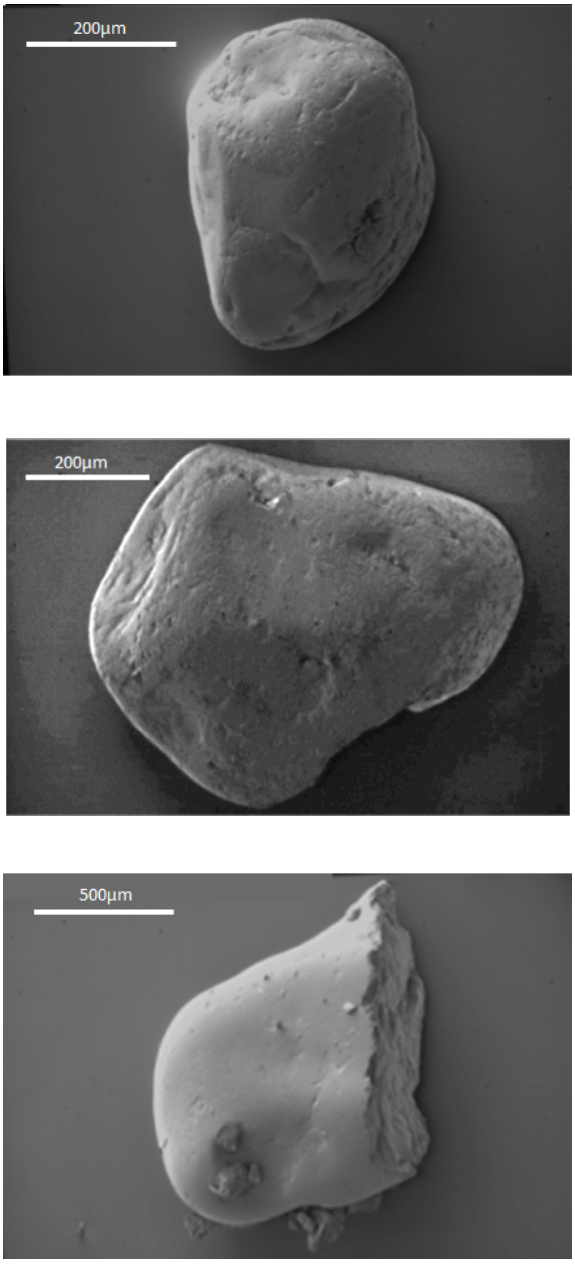
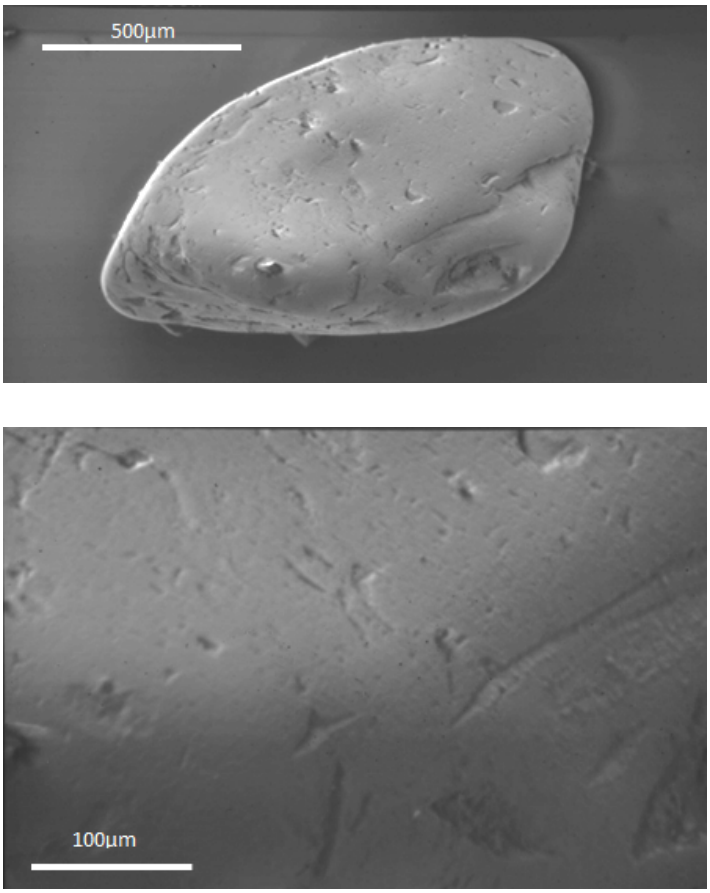


Figure 7.4: SEM images of an elongated grain (top) and of the observed fracture patterns (bottom)



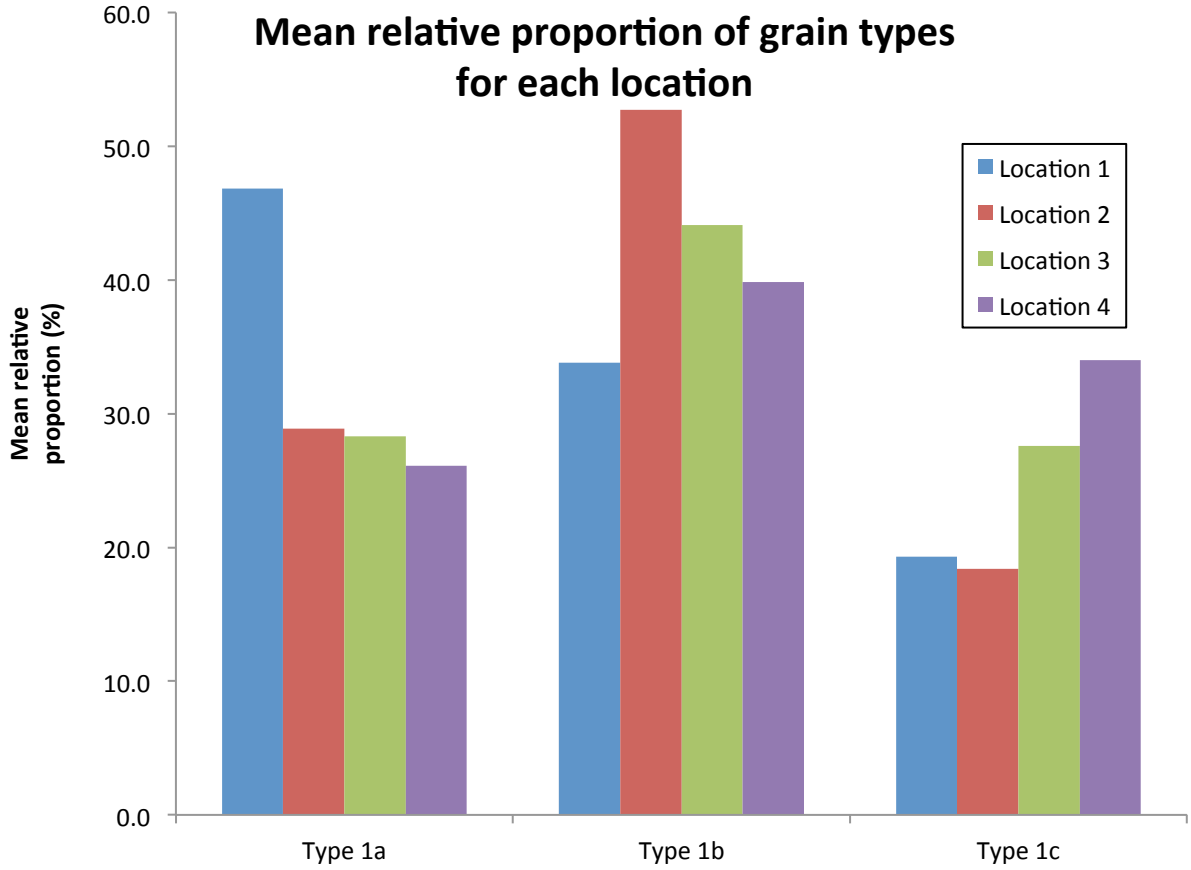
For each sampling location, the 50 grains were classified for each sample (n=5) to sub type and counted and the mean relative proportions for each grain type present was calculated for comparison of the four different locations. These data are displayed in table 7.4 and figure 7.5, respectively.

Table 7.4: Summary of the relative proportions of the three main grain types for each sample point.

Location	Sample Point	% Type 1a	% Type 1b	% Type 1c
1	A	74.3	14.3	11.4
	B	54.2	10.4	35.4
	C	34.0	48.0	18.0
	D	34.0	52.0	14.0
	E	37.8	44.4	17.8
	Range	34.0-74.0	10.4-52.0	11.4-35.4
	Mean	46.8	33.8	19.3
	%RSD	37.3	58.6	48.7
2	A	40.9	20.5	38.6
	B	36.0	42.0	22.0
	C	30.0	56.0	14.0
	D	12.0	80.0	8.0
	E	25.6	65.1	9.3
	Range	12.0-40.9	20.5-80.0	8.0-38.6
	Mean	28.9	52.7	18.4
	%RSD	38.4	43.1	68.4
3	A	15.4	15.4	69.2
	B	36.1	30.6	33.3
	C	16.0	78.0	6.0
	D	32.7	55.1	12.2
	E	41.5	41.5	17.1
	Range	15.4-41.5	15.4-78.0	6.0-69.2
	Mean	28.3	44.1	27.6
	%RSD	42.2	54.2	92.1
4	A	6.3	21.9	71.9
	B	31.0	12.1	56.9
	C	28.0	58.0	14.0
	D	42.9	44.9	12.2
	E	22.5	62.5	15.0
	Range	6.3-42.9	12.1-58.0	12.2-71.9
	Mean	26.1	39.9	34.0
	%RSD	51.2	55.6	83.1

Figure 7.5 shows the relative proportion of grains attributed to each grain type at each of the sample locations in Brockwell Park

Figure 7.5: Mean (n=5) relative proportion of grain types for each location



Each of the locations contained only three sub grain types; Ia rounded, Ib sub-rounded and Ic complete grain breakage. All of the samples analysed displayed a similar range and distribution of grain types, regardless of sample location. Grains from the samples from each location all displayed both weathering effects and surface markings consistent with having been formed in a high energy, aqueous environment. Furthermore, the distinctive, elongated grains observed were present at all locations. It has been demonstrated that across the UK, on average there are 2-3 distinct grain types present at a given location (137), so the lack of diversity within these samples is distinctive, but ultimately for this study does not provide additional discriminatory information for these samples taken from different locations within the park.

The large relative standard deviations (RSD) displayed in table 7.4 demonstrate the wide variation observed in the relative proportions of each grain type across the five sample points. For each grain type the intra-sample variation was such that, in most cases, the range of grain type ratios for each location were overlapping and the inter-location variability, determined by the differences between the mean relative proportions for each grain type at each location, was therefore far less than the differences between the five samples at each location. As a

result, it was not possible to discriminate samples from different locations with any degree of confidence based solely on the analysis of the surface textures of the quartz grains present.

The most notable differences between the locations were that the proportion of rounded grains (Ia) at Location 1 was higher than at the other locations, indeed the mean proportion of these grains at Location 1 was outside the range of proportions of rounded grains at the other locations. In addition, the mean proportion of sub-rounded grains (Ib) at Location 2 was outside the range of that grain type at Location 1. The relative proportions of each grain type are clearly dependent on one another therefore whilst these observations can be made it is not possible to use the proportion of a single grain type as a parameter for excluding different locations in isolation, since the parameters for the two other grain sub-types are inextricably correlated.

While quartz grain surface texture analysis has been able to provide valuable exclusionary, and in some cases diagnostic, intelligence and evidence in many forensic cases, the similarity of the underlying geology and the fluvial sediments that are widely present in the Thames Valley appear to have made the discrimination between sample locations within close proximity in this particular location difficult. This further highlights the importance of developing additional methods to add to the suite of techniques available for the analysis of geoforensic samples. It also therefore, demonstrates the value that HPLC analysis can have for the analysis of samples of close proximity where other arguably more well-established techniques are not able to yield discrimination between samples. It also serves as a valuable reminder of the importance of utilising a suite of multiple independent analytical techniques that can be applied to geoforensic samples on a case by case basis, in a manner that takes into account the different impinging variables of pertinence to each new case

7.3.2 Comparison of HPLC profiling with Wax Marker profiling by GC

In addition to comparing the performance of HPLC with an independent technique analysing the inorganic fraction of the samples, the performance of HPLC was also assessed relative to an established method for analysing the organic fraction of soils. The comparison of the HPLC results with the results obtained for the determination of wax markers by GC are presented below. The HPLC data has been presented in Chapter 5 (section 5.3.2) and in Chapter 6 sections (6.3.1, 6.3.2, 6.3.3 and 6.3.4 for the Edinburgh, Aberdeen, London, and New York City sites, respectively, at the Winter time point only), however a summary is provided here to enable comparison with the results of the GC analysis.

Figure 7.6: HPLC Profiles from Lochend Park, Edinburgh

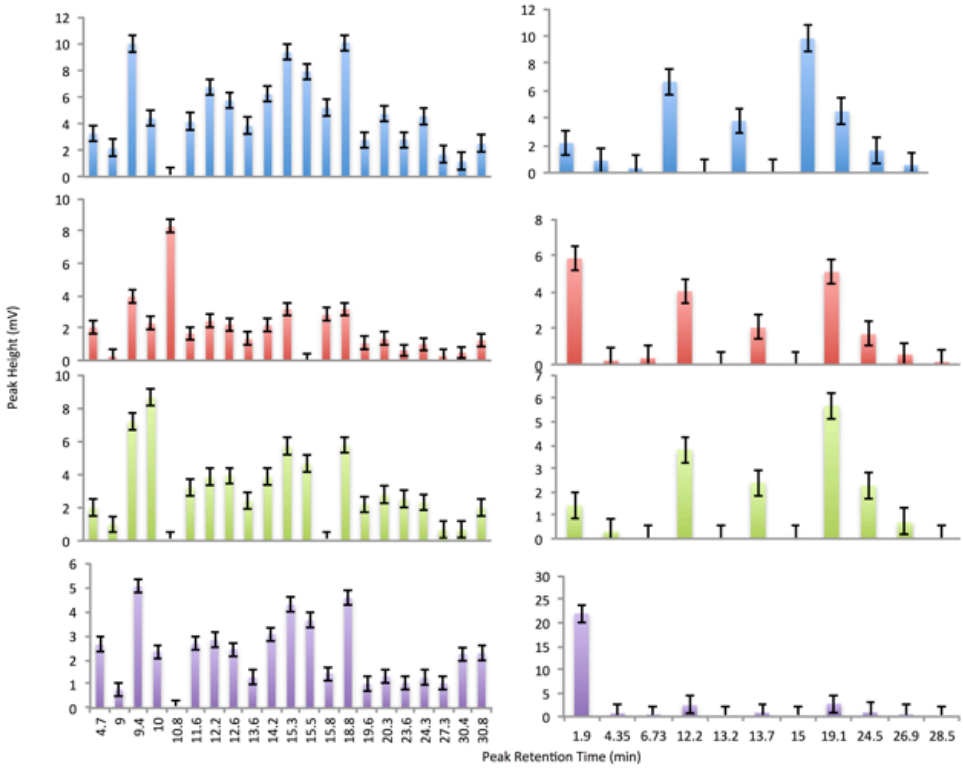


Figure 7.6: Edinburgh HPLC Profiles using marker set A (left) and B (right) for managed grassland (blue), adjacent to fresh water (red), unmanaged land (green) and woodland (purple).

It was possible to discriminate all four locations within Lochend Park, Edinburgh on the basis of the profiles of HPLC peak set A (Figure 7.6), as presented in sections 5.3.2 and 6.3.1 and by McCulloch et al. (34). The presence of the peak at 10.8min and the absence of the peak at 15.8min were useful markers for soils adjacent to fresh water and unmanaged land, respectively, while the profiles of woodland and managed grassland were more similar but could be separated by differences in the relative size of certain pairs of peaks.

The profiles for HPLC peak set B (Figure 7.6) were not as easily distinguishable at each of the four locations in Lochend Park, Edinburgh. Under scrutiny however, each profile was different from the others (34). The woodland samples were distinctive in that they had large peaks at 1.9min, while soils adjacent to fresh water were distinct in having their two largest peaks at 1.9 and 19.1min, that were similar in size to one another. It was more difficult to visually discriminate the profiles of managed grassland and unmanaged land, however, as discussed in McCulloch et. al. (34) the small peaks present in the managed grassland profiles at 6.7 and 28.5min were absent in the samples from unmanaged land.

Figure 7.7: Wax Marker Profiles from Lochend Park, Edinburgh

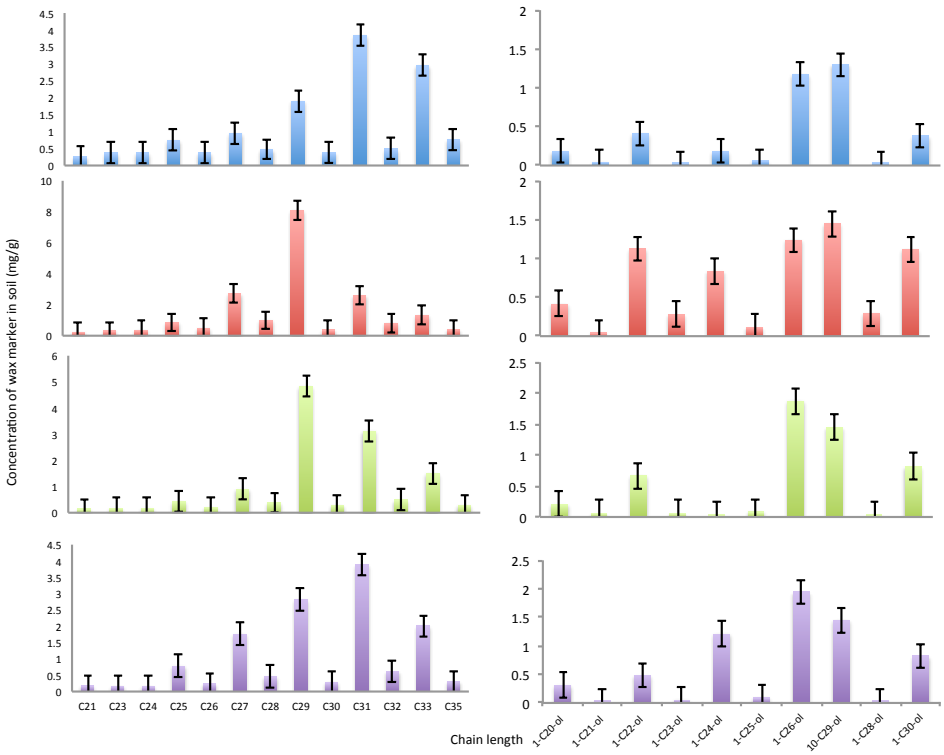


Figure 7.7: Edinburgh Profiles of n-alkanes (left) and alcohols (right) for managed grassland (blue), adjacent to fresh water (red), unmanaged land (green) and woodland (purple).

The wax marker profiles (Figure 7.7) allowed for clear visual discrimination of the four sample locations at Lochend Park. The location adjacent to fresh water and unmanaged land were distinct in that C29 was the most abundant marker, and could be separated from one another, visually, through the presence of 1-C24-ol for the soils adjacent to fresh water, a marker that was absent in unmanaged land. The increase in concentration for the series C25, C27, C29 and C31 was characteristic of managed grassland and woodland soils, however the increase appeared more linear for woodland soils and exponential for the managed grassland, these two locations could also be distinguished from one another through the ratio of 1-C24-ol to 1-C26-ol, which was much higher for woodland soils.

Table 7.5: CDFA Results for Lochend Park, Edinburgh

Marker Type	Classification Accuracy %	Wilks' Lambda Significance test of functions			% Variance Function 1	% Variance Function 3	% Variance Function 2
Edinburgh		1-3	2-3	3			
HPLC Set A	100.0	.000	.000	.022	88.9	8.3	2.7
HPLC Set B	100.0	.000	.018	.397	62.4	32.5	5.0
Wax Markers	100.0	.000	.000	.025	70.8	21.3	7.9

At Lochend Park, all three sets of markers gave 100% accuracy in grouping samples to the correct location when used in CDFA (Table 7.5). The functions created for these sets of variables correctly predicted the location that each sample belonged to. The first function accounted for the greatest amount of variation between the groups for all three sets of markers, at 88.9%, 62.4% and 70.8% of the observed variance for HPLC set A, HPLC set B and the Wax Markers, respectively. When either all three Functions, the group differences were statistically significant at the >99% confidence interval.

Figure 7.8: CDFA Scatter Plots for Lochend Park, Edinburgh

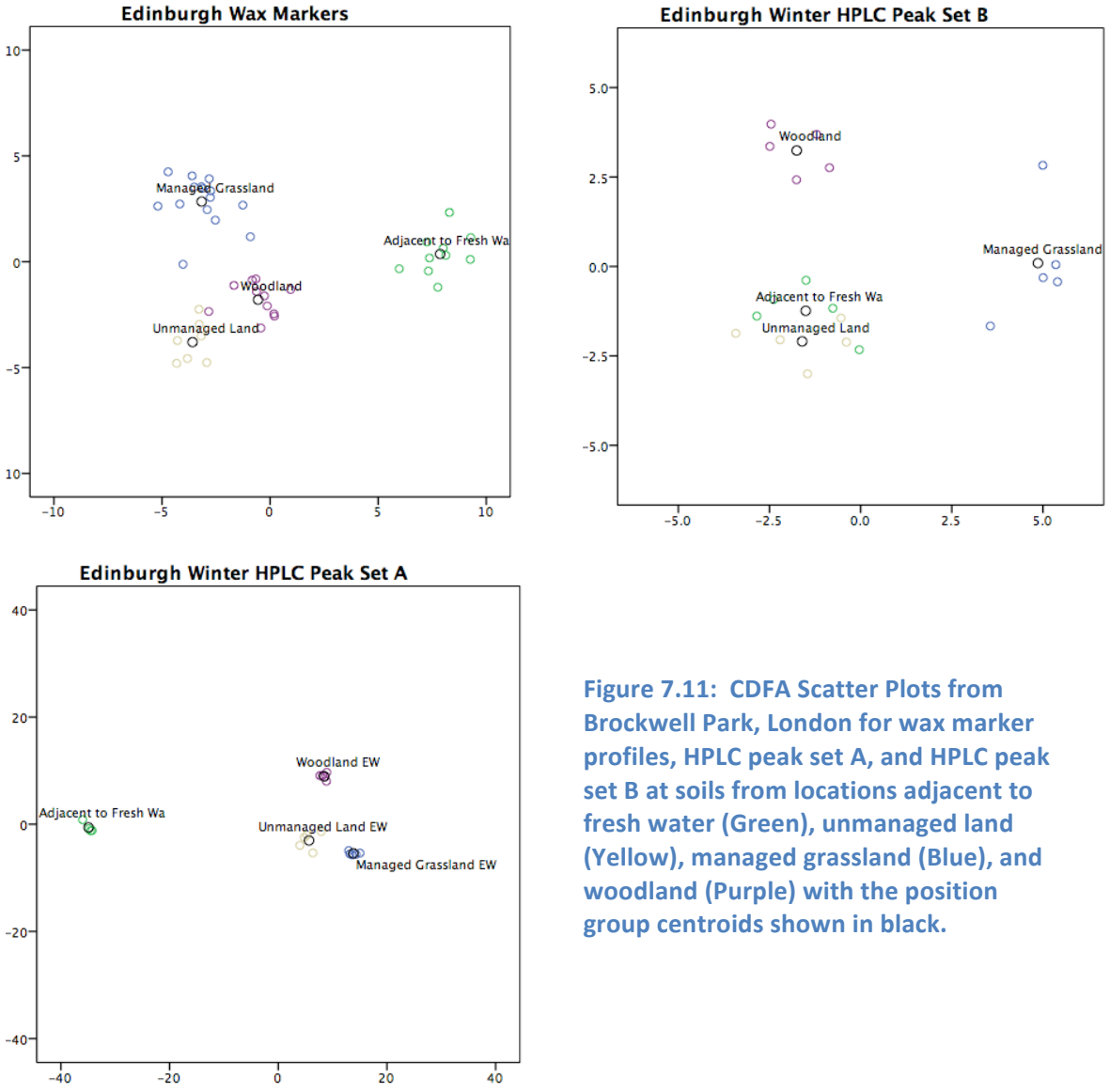


Figure 7.11: CDFA Scatter Plots from Brockwell Park, London for wax marker profiles, HPLC peak set A, and HPLC peak set B at soils from locations adjacent to fresh water (Green), unmanaged land (Yellow), managed grassland (Blue), and woodland (Purple) with the position group centroids shown in black.

The scatter plots for function 1 and 2 (Figure 7.8) show sample groups clustering according to their sample location, with clear separation between the groups evident for all three sets of markers. All samples were grouped correctly, as full separation of groups is achieved with function 3, which is not displayed on these plots.

Figure 7.9: HPLC Profiles from Brockwell Park, London.

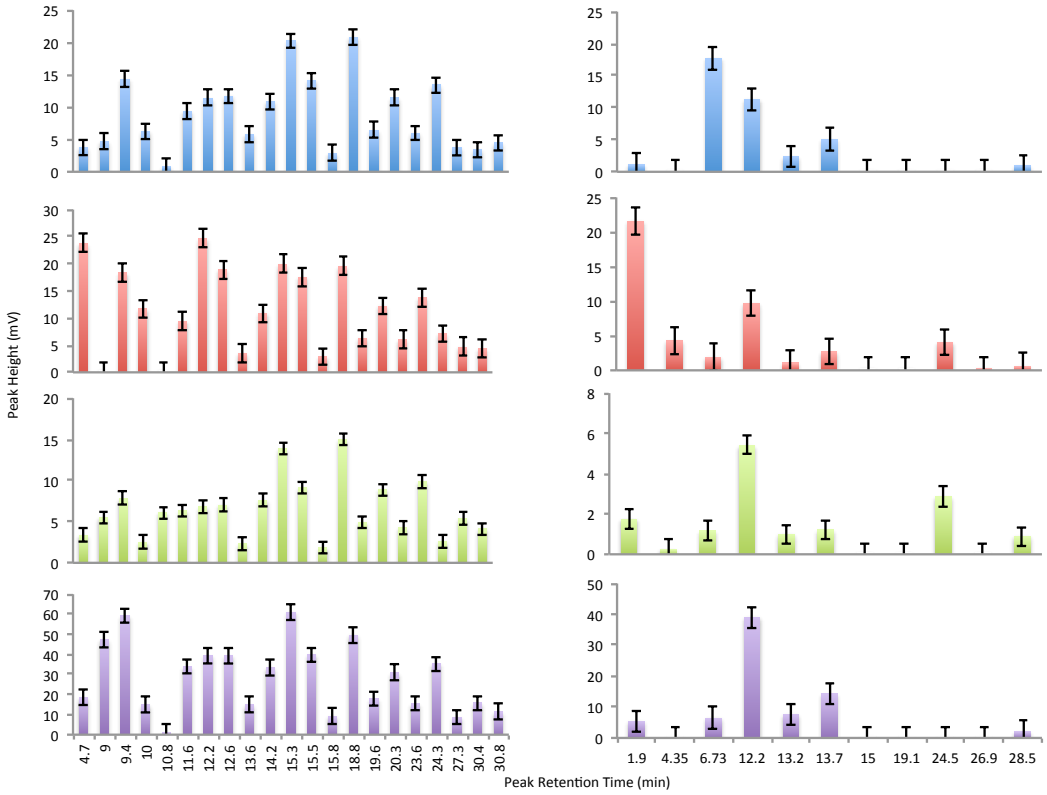


Figure 7.9: London HPLC Profiles using marker set A (left) and B (right) for managed grassland (blue), adjacent to fresh water (red), unmanaged land (green) and woodland (purple).

All four locations within Brockwell Park, London (Figure 7.9) could be distinguished by the profiles of HPLC peak set A. Comparison of the two tallest peaks for each location was useful in discriminating the samples in this dataset. For soils adjacent to fresh water the two tallest peaks were 4.7 and 12.2min, whereas for woodland soils the peaks at 9.4 and 15.3min were largest, on the other hand the peaks at 15.3 and 18.8min were the largest peak in the profiles of both managed grassland and unmanaged land, suggesting the possibility that these peaks are indicative of grassland soils. Peak height ratios were also used to distinguish the sites, and full discussion of the visual similarities and differences is provided in sections 5.3.2 and 6.3.3, and in McCulloch et. al. (34)

The HPLC profiles for peak set B (Figure 7.9) allowed all four samples locations within Brockwell Park, London to be discriminated visually. To summarise the findings discussed in sections 5.3.2 and 6.3.3, and in McCulloch et. al. (34) the large size of the peak at 6.73min relative to the peak at 12.2min distinguishes managed grassland from all other sample locations, while the large size of the peaks at 1.9min compared to all other peaks is distinctive of the profiles in soils adjacent to fresh water. The profiles of soils from unmanaged land and

woodland were visually more similar, however can be separated due to the presence of the peak at 24.5min for unmanaged land which is absent in woodland samples.

Figure 7.10: Wax Marker Profiles from Brockwell Park, London

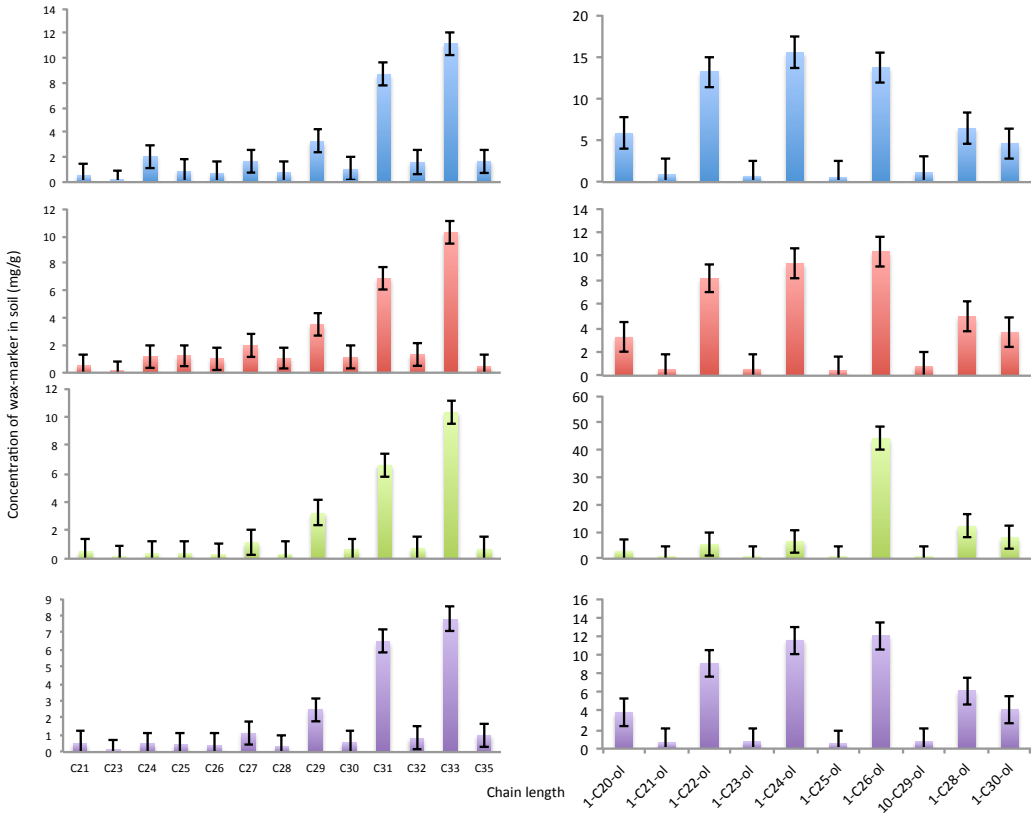


Figure 7.10: London Profiles of n-alkanes (left) and alcohols (right) for managed grassland (blue), adjacent to fresh water (red), unmanaged land (green) and woodland (purple).

The different locations in Brockwell Park were not so easily discriminated using their wax marker profiles (Figure 7.10) and only the profile from unmanaged land was visually distinct from the others with an exceptionally high concentration of 1-C26-ol at 44 mg/g compared to approximately 10mg/g for the other locations. More subtle differences in the relative concentrations of 1-C24-ol and 1-C26-ol allowed tentative discrimination of managed grassland from woodland and soils adjacent to fresh water, as the 1-C24-ol peak was 13% larger than the 1-C26-ol for the former, while it was 4% and 9% smaller than 1-C26-ol for woodland and soils adjacent to fresh water, respectively. Likewise, small differences were observed in the size difference between C31 and C33 at these two locations with an increase of 20% and 49% from C31 to C33 for woodland and soil adjacent to fresh water. It should be noted that these more subtle differences were fairly small in magnitude compared to the variability in the results at each location, as indicated by the error bars.

Table 7.6: CDFA Results for Brockwell Park, London

Marker Type	Classification Accuracy %	Wilks' Lambda Significance test of functions			% Variance Function 1	% Variance Function 2	% Variance Function 3
London		1-3	2-3	3			
HPLC Set A	100.0	0.000	0.002	0.034	89.7	7.0	3.3
HPLC Set B	90.0	0.000	0.041	0.684	84.8	13.9	1.3
Wax Markers	80.4	0.000	0.011	0.119	81.7	11.6	6.6

At Brockwell Park (Table 7.6), only HPLC set A gave 100% accuracy when used in CDFA, while 90% and 80.4% of samples were assigned to the correct group for HPLC set B and the wax markers, respectively. The first function accounted for 89.7%, 84.8% and 81.7% of the observed variance for HPLC set A, HPLC set B and the wax markers, respectively. When all three Functions were used, the group differences were statistically significant at the >99% confidence interval.

Figure 7.11: CDFA Scatter Plots for Brockwell Park, London

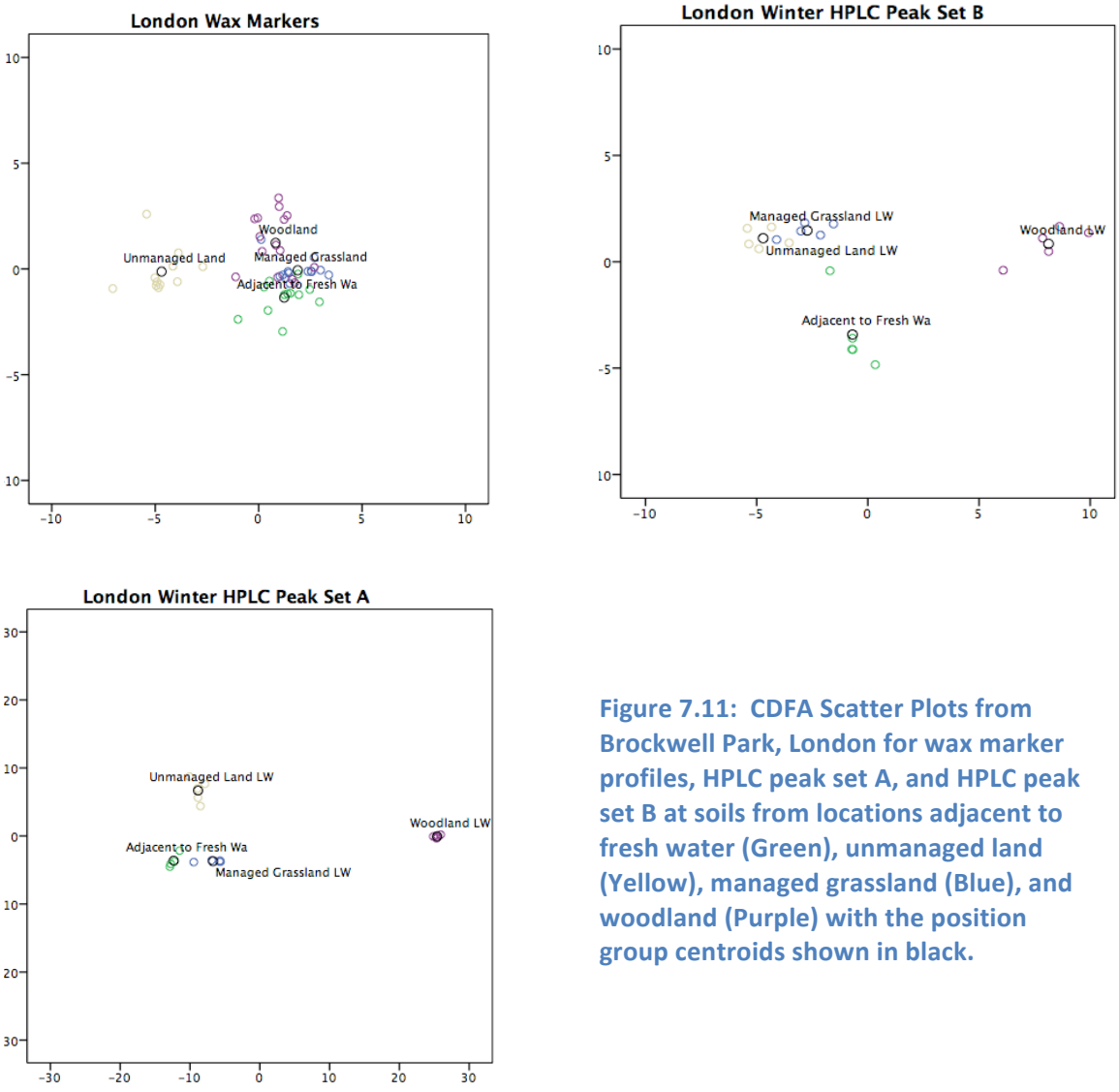


Figure 7.11: CDFA Scatter Plots from Brockwell Park, London for wax marker profiles, HPLC peak set A, and HPLC peak set B at soils from locations adjacent to fresh water (Green), unmanaged land (Yellow), managed grassland (Blue), and woodland (Purple) with the position group centroids shown in black.

The inaccuracies in the wax marker groupings at Brockwell Park were highest for the managed grassland, with four samples misclassified as adjacent to fresh water and one sample attributed to the woodland location. Woodland samples were also highly misclassified, one was assigned to managed grassland while a further four were grouped with soil adjacent to fresh water. Two samples were misclassified for HPLC set B, one sample from managed grassland was predicted to belong to the unmanaged land group, while one sample from the location adjacent to fresh water was incorrectly assigned to the managed grassland soil group (Figure 7.11).

Figure 7.12: HPLC Profiles from Central Park, New York.

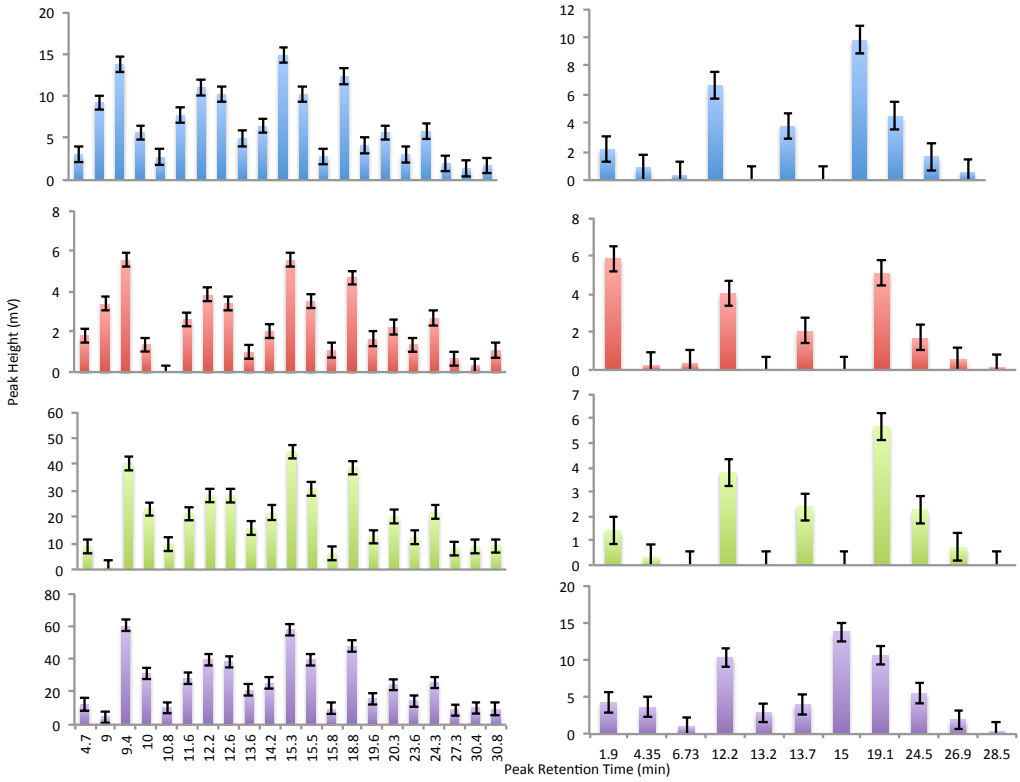


Figure 7.12: New York HPLC Profiles using marker set A (left) and B (right) for managed grassland (blue), adjacent to fresh water (red), unmanaged land (green) and woodland (purple).

The results for Central Park, New York City are discussed fully in sections 5.3.2 and 6.3.4, and in McCulloch et. al. (34). All four locations could be discriminated on the basis of their HPLC profiles for peak set A (Figure 7.12). The location adjacent to fresh water could be unambiguously distinguished through the absence of a peak at 10.8min, while the absence of the peak at 9min was a unique feature of the soils from unmanaged land at this site. The profiles of the managed grassland and woodland locations were very similar, however there was a noticeable difference between the two locations in the size of the peak at 9min compared to its neighbour at 4.7min.

The profiles obtained for peak set B also varied across the four locations in Central Park (Figure 7.12), New York City. Soil profiles for unmanaged land could be separated from the other three locations by the absence of the peak at 1.9min, while the profiles from soil adjacent to fresh water were most noticably different from the other locations in the ratio of the peak at 1.9min compared to the peak at 12.2 min. The peaks in the managed grassland samples were generally three times as large as those for the location adjacent to fresh water, while woodland samples were approximately twice the size obtained for managed grassland Again,

the relative sizes of the peaks, and peak ratios discussed in sections 5.3.2 and 6.3.4, and in McCulloch et. al. (34) provided additional ways to separate the groups visually.

Figure 7.13: Wax Marker Profiles from Central Park, New York

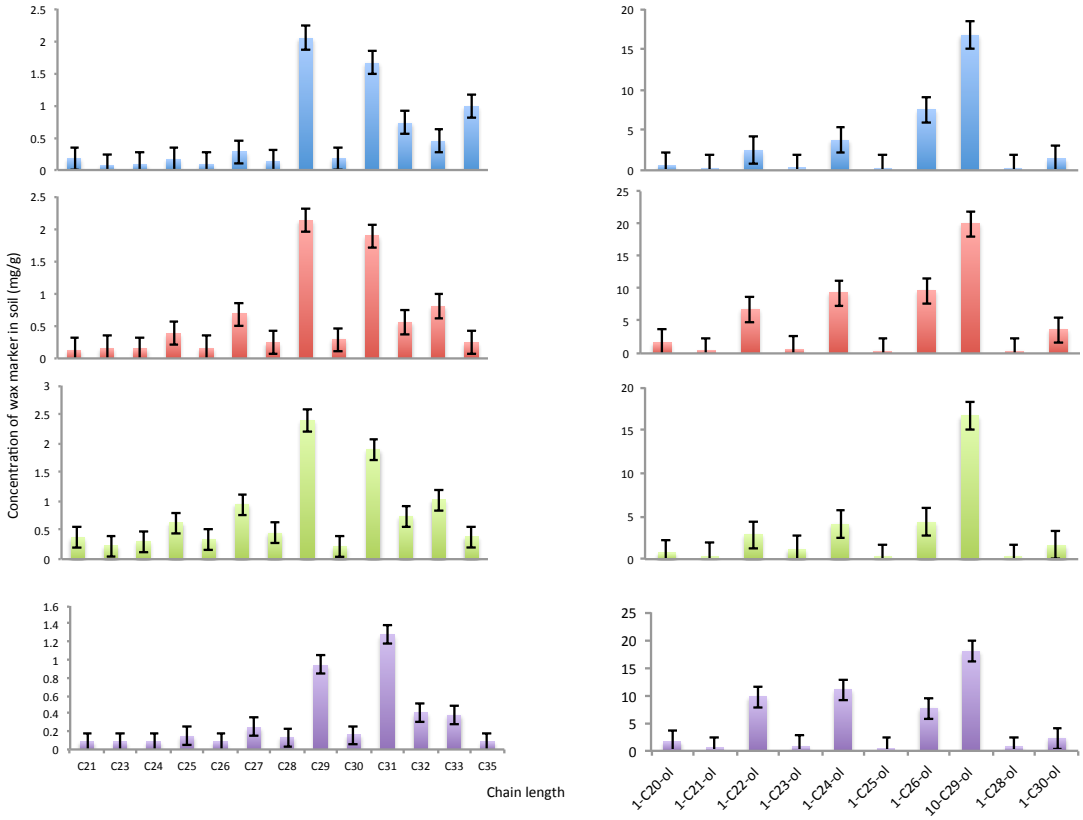


Figure 7.13: New York Profiles of n-alkanes (left) and alcohols (right) for managed grassland (blue), adjacent to fresh water (red), unmanaged land (green) and woodland (purple).

The wax markers profiles (Figure 7.13) allow woodland soils to be easily discriminated since C31 was present at a higher concentration than C29, whereas C29 was more abundant than C31 for the other locations. Managed grassland contained elevated levels of C35 compared to C33, whereas for the other locations, C35 was less abundant than C33. The most prominent feature distinguishing the unmanaged land from soils adjacent to fresh water was the ratio of 10-C29-ol to 1-C26-ol which was much higher for the former, at 3.9:1, compared to 2:1 for the latter. While it was possible to visually discriminate the four profiles, many of the distinguishing features were small in magnitude compare to the variability of the data within each location.

Table 7.7: CDFA Results for Central Park, New York City

	Classification Accuracy %	Wilks' Lambda Significance test of functions			% Variance Function 1	% Variance Function 2	% Variance Function 3
New York		1-3	2-3	3			
HPLC Set A	100.0	0.000	0.000	0.005	92.3	5.8	1.9
HPLC Set B	100.0	0.000	0.000	0.071	73.3	24.1	2.5
Wax Markers	89.9	0.000	0.000	0.003	66.1	25.3	8.6

Both HPLC sets A and B gave 100% grouping accuracy at the New York site, while the wax markers correctly assigned 89.9% of the samples to their groups (Table 7.7). The differences between groups were statistically significant for all marker sets at the >99% confidence level with functions 1-3. Function 1 accounted for 92.3%, 73.3% and 66.1% of the variation between groups for HPLC sets A and B, and the wax markers, respectively.

Figure 7.14: CDFA Scatter Plots for Central Park, New York

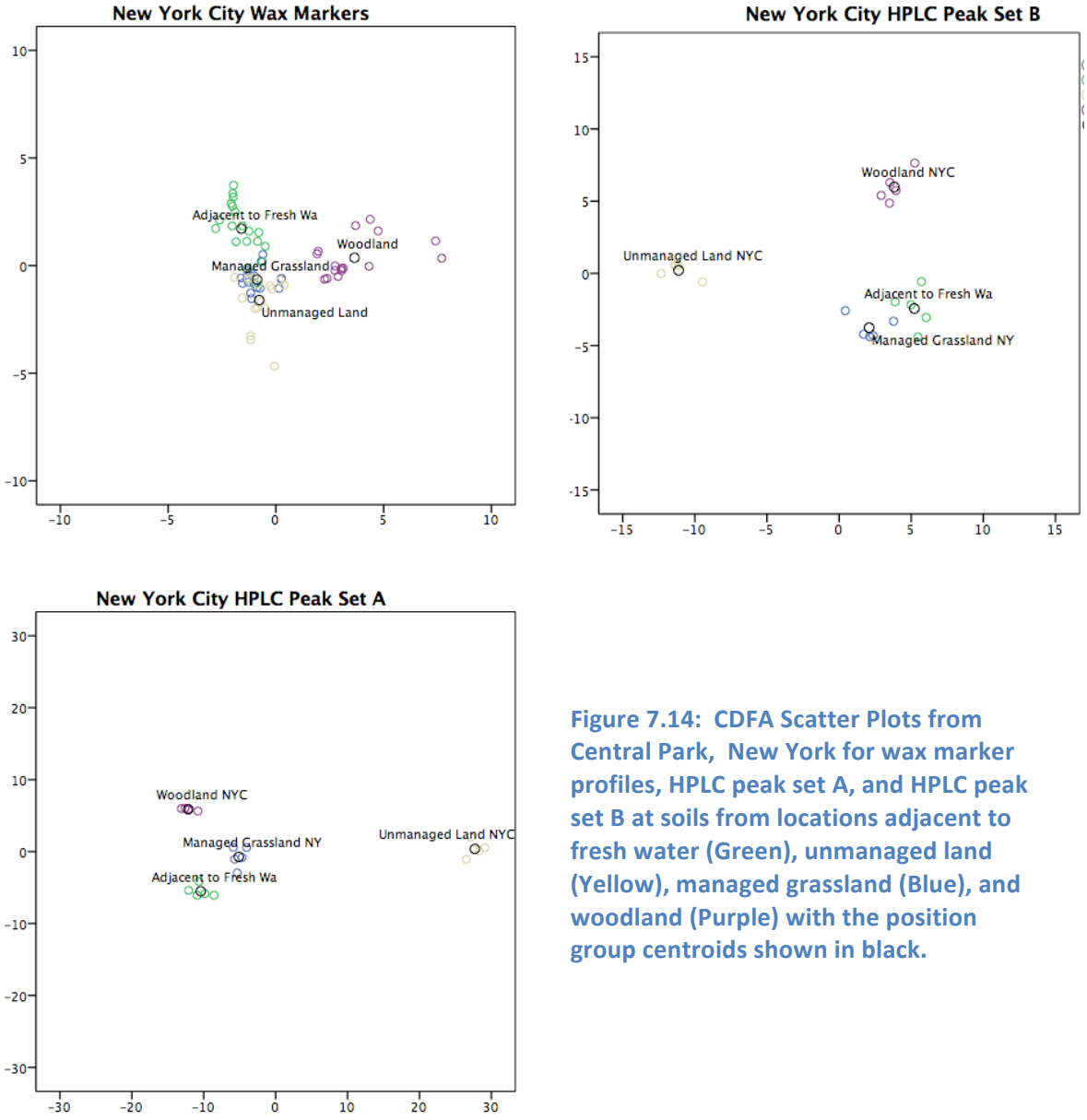


Figure 7.14: CDFA Scatter Plots from Central Park, New York for wax marker profiles, HPLC peak set A, and HPLC peak set B at soils from locations adjacent to fresh water (Green), unmanaged land (Yellow), managed grassland (Blue), and woodland (Purple) with the position group centroids shown in black.

The accuracy of the groupings at the Central Park site are reflected in the CDFA scatter plots (Figure 7.14), with good spacing between groups evident for HPLC peak sets A and B, and some overlap of sample groups for the wax markers. CDFA on the wax marker data misclassified one sample from managed grassland as unmanaged land, two samples from the location adjacent to fresh water were assigned to managed grassland and one inaccurately placed in the unmanaged land group, while three samples from unmanaged land were attributed to unmanaged land, however all the woodland samples were grouped accurately.

Figure 7.15: HPLC Profiles from Craigiebuckler, Aberdeen.

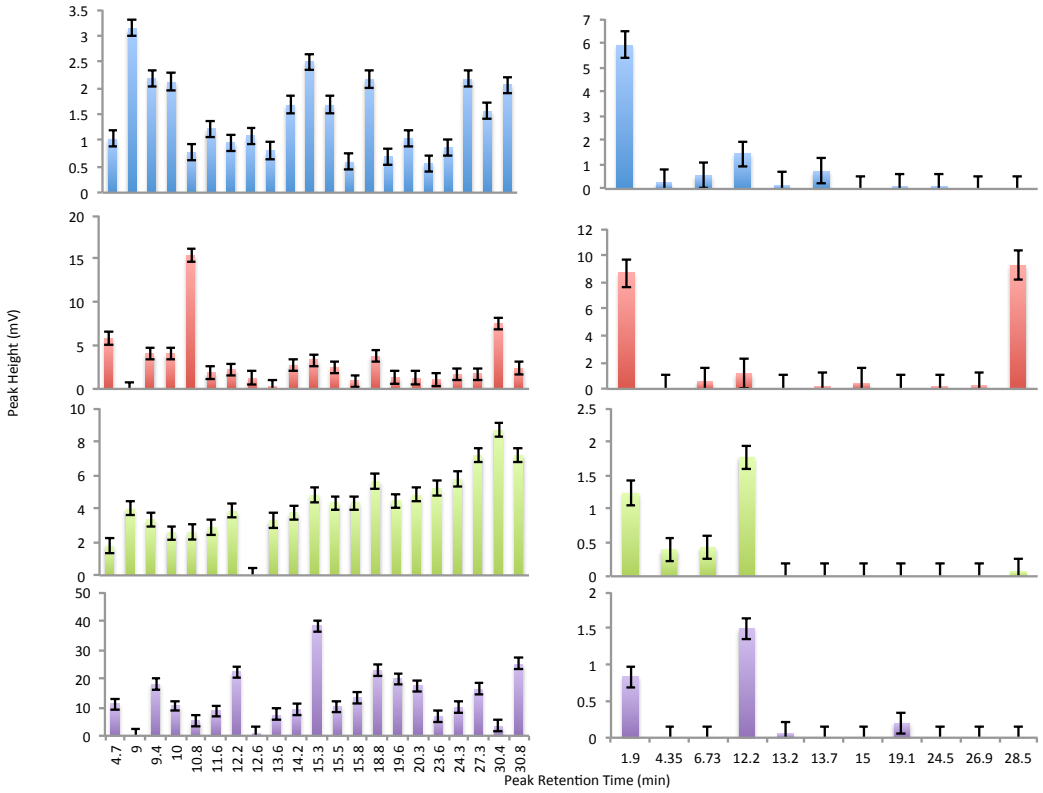


Figure 7.15: Aberdeen HPLC Profiles using marker set A (left) and B (right) for managed grassland (blue), adjacent to fresh water (red), unmanaged land (green) and woodland (purple).

The profiles for HPLC peak set A allowed each of the four locations in Craigiebuckler Estate, Aberdeen to be distinguished visually (Figure 7.15). As outlined sections 5.3.2 and 6.3.2, and in McCulloch et. al. (34), for managed grassland samples were distinguished by having their largest peak at 9.4min and the profiles of soils adjacent to fresh water were unique in having their largest peak at 10.8min. The large relative height of the peak at 30.8min was distinctive of woodland soil profiles, and the profile of the samples from unmanaged land was distinctive with the highest peak at 30.4min.

Using peak set B (Figure 7.15) the samples from the four locations within the Craigiebuckler Estate, Aberdeen were easily distinguished by their HPLC profiles. Full discussion of the differences between locations are provided in sections 5.3.2 and 6.3.2, and in McCulloch et. al. (34), however comparison of the retention time of and ratio between the two largest peaks at each location was useful in grouping the samples from managed grassland and those from soils adjacent to fresh water, while the ratio between the 1.9min and 12.2min peak pairs separated woodland soils from unmanaged land.

Figure 7.16: Wax Marker Profiles from Craigiebuckler, Aberdeen

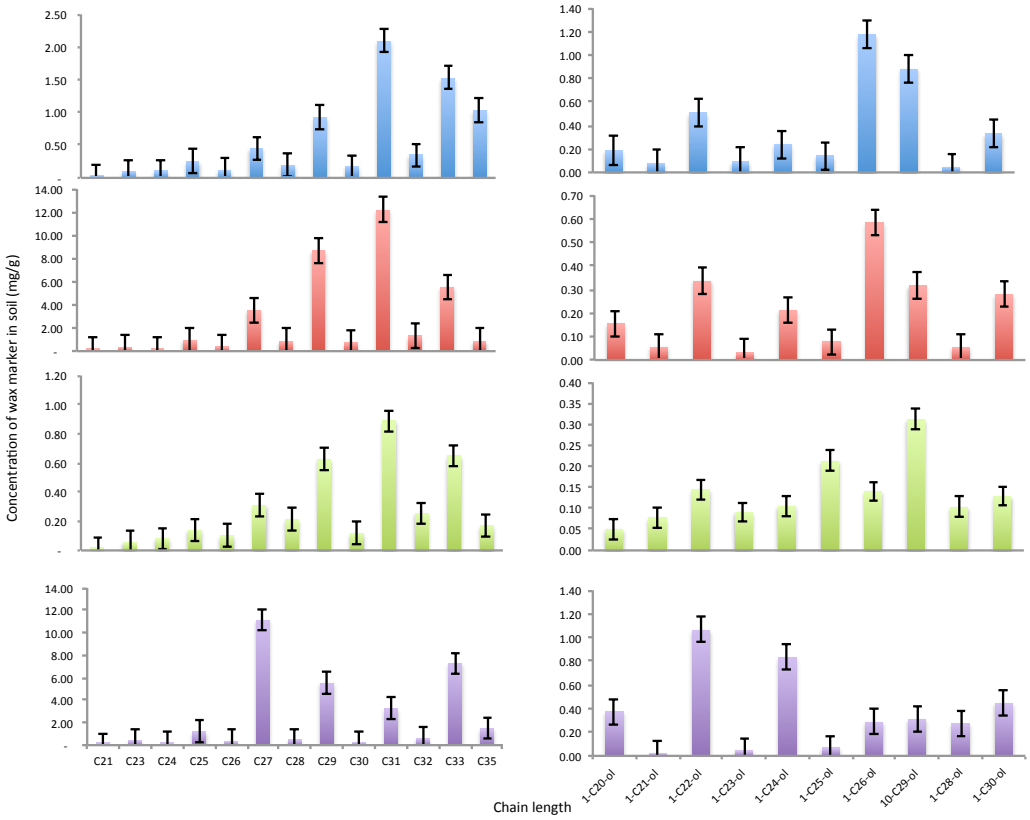


Figure 7.16: Aberdeen Profiles of n-alkanes (left) and alcohols (right) for managed grassland (blue), adjacent to fresh water (red), unmanaged land (green) and woodland (purple).

Woodland soils at the Craigiebuckler site could be easily identified from their wax marker profile (Figure 7.16), since C27 was the most abundant *n*-alkane, whereas C31 had the highest concentration of the *n*-alkanes for the other locations. Unmanaged land was distinct from the other locations since 10-C29-ol was the most concentrated of the alcohols. The remaining locations, adjacent to fresh water and managed grassland, could be differentiated by the ratios of C33:C35, which were 1.5:1 and 6.4:1, respectively.

Table 7.8: CDFA Results for Craigiebuckler, Aberdeen

Marker Type	Classification Accuracy %	Wilks' Lambda Significance test of functions			% Variance Function 1	% Variance Function 3	% Variance Function 2
Aberdeen		1-3	2-3	3			
HPLC Set A	94.7	.001	.147	.531	90.6	7.8	1.5
HPLC Set B	100.0	.000	.000	.014	97.4	2.4	0.2
Wax Markers	86.2	.000	.000	.397	68.0	26.6	5.4

At the Aberdeen site (Table 7.8), 94.7% of samples were grouped correctly using HPLC set A, while HPLC set B provided 100% accuracy and the wax markers predicted the correct group in 86.2% of cases. The differences between groups were statistically significant at the >99% confidence interval when functions 1-3 were used for all marker sets.

Figure 7.17: CDFA Scatter Plots for Craigiebuckler, Aberdeen

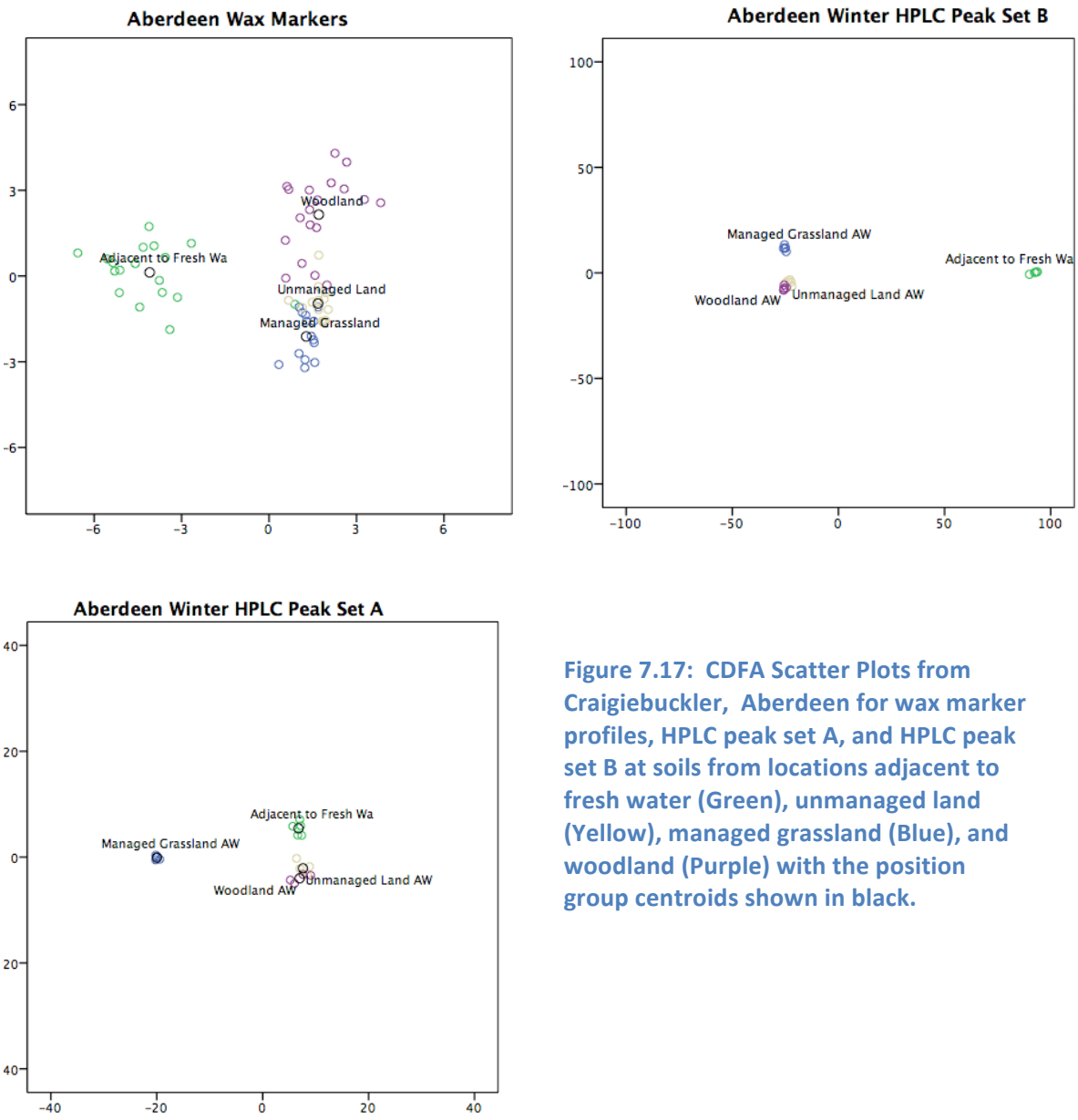


Figure 7.17: CDFA Scatter Plots from Craigiebuckler, Aberdeen for wax marker profiles, HPLC peak set A, and HPLC peak set B at soils from locations adjacent to fresh water (Green), unmanaged land (Yellow), managed grassland (Blue), and woodland (Purple) with the position group centroids shown in black.

Using HPLC Peak Set A, all but one sample was correctly classified across all four sites. One sample from the unmanaged location in Craigiebuckler Estate, Aberdeen was misclassified as having originated from the woodland location. All samples were correctly assigned for HPLC set B. Three samples taken from managed grassland were attributed to unmanaged land using the wax marker data, in addition one sample from the location adjacent to fresh water was misclassified as unmanaged land, while one sample from unmanaged land was grouped with

woodland soils, and four woodland samples were incorrectly classified as unmanaged land (Figure 7.17).

When either of the HPLC marker sets or wax markers were unable to give 100% accuracy in groupings, further CDFA were performed using the results of both the GC and HPLC results in conjunction with one another (Table 7.9). Using only wax markers, the accuracy rates were 86.2%, 89.9% and 80.4& for the Aberdeen, New York City, and London sites, using only HPLC set A the accuracy rate was 94.7% in Aberdeen, and for HPLC set B in London, the accuracy was 90%. Using both the HPLC and wax markers in the same CDFA, the following results were observed.

Table 7.9: CDFA Results for the combined use of HPLC and Wax Markers.

Marker Type	Classification Accuracy %	Wilks' Lambda Significance test of functions			% Variance Function 1	% Variance Function 2	% Variance Function 3
London Results for Wax and HPLC Markers		1-3	2-3	3			
Set A	100.0	0.000	0.002	0.034	89.7	7.0	3.3
Set B	100.0	0.000	0.019	0.246	82.7	14.2	3.1
Aberdeen Results for Wax and HPLC Markers							
Set A	100.0	0.000	0.000	0.000	93.0	4.8	2.2
Set B	100.0	0.000	0.000	0.046	98.6	1.3	0.2
New York City Results for Wax and HPLC Markers							
Set A	100.0	0.000	0.005	0.084	93.9	4.6	1.5
Set B	100.0	0.000	0.027	0.581	95.1	4.6	0.3

In all of the cases where less than perfect accuracy was achieved using only the HPLC markers or only the wax markers, 100% accuracy was achieved using the two types of marker to complement one another (Table 7.9). Furthermore, the separation between groupings was statistically significant at the 99% confidence level. At the London site, functions 1-3 explained 89.7%, 7.0% and 3.3%, respectively, of the variation between the groups using the wax markers and HPLC set A, and 82.7%, 14.2% and 3.1% of the variance, using HPLC set B. For the Aberdeen site, the first three canonical functions explained 93.0%, 4.8% and 2.2% of the variance using HPLC set A to complement the wax markers, and 98.6%, 1.3% and 0.2% using HPLC set B. The CDFA results for these data sets are displayed in figure 7.18.

Figure 7.18: CDFA scatter plots for the combined use of HPLC and Wax Markers.

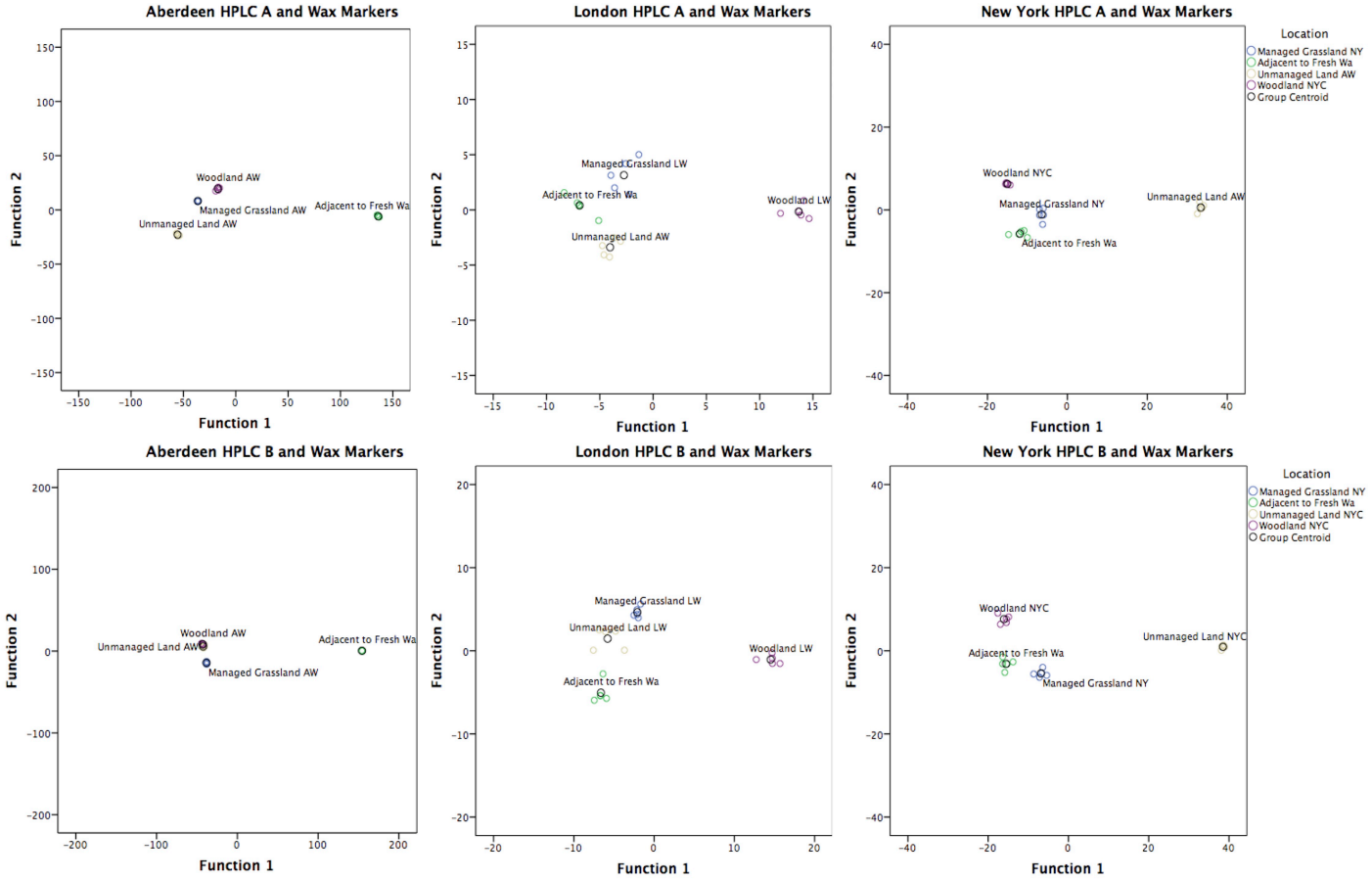


Figure 7.18: CDFA Scatter Plots for the combined use of wax marker profiles and HPLC peak sets, at sites where either marker set could not provide 100% accuracy, for soils from locations adjacent to fresh water (Green), unmanaged land (Yellow), managed grassland (Blue), and woodland (Purple) with the position group centroids shown in black.

Both the HPLC and GC profiling techniques provided strong results, with high accuracy rates when discriminated using CDFA. The statistical analysis shows HPLC to offer exceptionally high accuracy rates, which are discussed fully in Chapter 5, section 5.3.2 and Chapter 6, sections 6.3.1-6.3.4, and in McCulloch et. al (34). Accuracy rates were slightly poorer for the GC markers than either of the HPLC marker sets, with the exception of Lochend Park where all three sets of markers gave perfect accuracy. The two HPLC marker sets were able to improve the accuracy achievable using only the wax markers, demonstrating that HPLC can add value to geo-forensic investigations when used alongside established techniques. Interestingly, in the two cases where neither set of HPLC markers gave 100% accuracy when used on their own, for example HPLC set B for London and HPLC set A for Aberdeen, addition of the wax markers, which offered poorer accuracy than either HPLC set in the initial analyses, also improved the accuracy of discrimination to 100%.

Visual comparison of the profiles also allowed for discrimination between the four locations at each site, however the visual differences were generally more distinct for the HPLC data, making this process quicker and easier for these data sets. Furthermore, with more points for comparison, the degree of certainty with which a visual assessment could be made was greater for the HPLC, however for all sites the GC profiles allowed for at least one of the locations to be unambiguously identified visually. Visual assessments of the profiles is naturally an extremely subjective process, and the ease with which judgments on group membership can be made is likely to differ from analyst to analyst, as is the case in other domains within forensic science. Furthermore, the variability in the data within each location was such that the subtle differences in peak ratios that facilitated complete discrimination of all four locations, and were particularly useful for the wax markers, should be treated with caution when interpreting evidence, as these small differences may not be reliably detected across all sample replicates. This high internal variability presents an additional challenge when sample quantities are limited such that the number of sample replicates is restricted. In this regard, wax marker profiling offers a significant advantage over HPLC, since the analysis can be performed on sample quantities as low as 13mg (111).

The high variability observed in the data is likely to be a reflection of the heterogeneous nature of soil, however there are a number of potential sources of variability in the methodology which must be considered, for both the HPLC and GC techniques. The sample preparation technique for the GC analysis was far more complex than for the HPLC, with multiple concentration and reconstitution steps, using very small volumes of 20-60µl sample. It is more difficult to achieve analytical precision and accuracy when working with smaller volumes of

sample since small systematic errors have a larger relative effect on the results. Furthermore, the large number of steps in the analytical method presented multiple opportunities for small human errors, such as inadvertent, inaccurate weighing or imprecise pipetting, to accumulate to create comparatively large variability in the resulting data.

The HPLC sample preparation technique was simpler than the GC, with only one weighing and dilution step, a batch of samples could be prepared within a two hours compared to five days for the GC. Aside from the clear practical and financial advantages of a simpler sample preparation method, the reduced opportunity for analyst variability suggests that the error bars evident in the HPLC data arise from true differences in the samples, or from the limitations of the instrument parameters chosen during method development. The HPLC methodology did not include a grinding step, as the technique was developed with the aim of retaining as much of the innate variation between samples as possible, so that rare and informative peaks were not diluted out through homogenisation. It is therefore recommended that, the effect of incorporating a grinding step should be investigated to establish whether this could improve the representativeness of each sample point or maximise the extraction efficiency thereby allowing better sensitivity and the potential to reduce sample quantities.

7.4 Conclusions

The results of this study confirm that HPLC is a highly accurate and practical technique for the analysis of geoforensic evidence, and that it is applicable for discrimination of close proximity sites across a range of geographical locations. In contrast, quartz grain surface texture analysis, which is an established geoforensic technique that has provided useful information to many criminal investigations in the past, was unable to provide the same level of discrimination at this location. The wax marker profiles performed well across all four sites, however the accuracy rates achieved with CDFA were better for HPLC, and in visual comparison, the wax marker profiles were less easily discernable than for HPLC. HPLC has, therefore, been demonstrated to be of value in situations where the more traditional, independent analytical technique would not have been able to provide exclusionary results for samples of differing provenance at this local scale. It is also significant that the HPLC performs well when compared to the GC technique, which has been validated to industry standards and is an accepted technique within geoforensic casework.

The HPLC profiles were useful not only when combined with CDFA, but also when compared visually. In addition to offering more visually distinct profiles to both the QGSTA and GC, and improved accuracy in the statistical analysis compared to the GC technique, the HPLC analysis

is significantly quicker, cheaper and simpler to perform than the GC technique, which reduces error rates and improves the confidence with which conclusions can be made when comparing samples. The cost, speed and simplicity of the HPLC sample preparation was comparable to the QGSTA sample preparation, however the process of visual analysis and quartz grain counting required for QGSTA was far more labour intensive, and therefore costly, than using the pre-selected HPLC marker sets as variables for CDFA, since software automated peak integration is now possible.

Whilst there are clear practical benefits to using the HPLC technique, due to the greater complexity of the sample preparation in the GC analysis, wax marker analysis has been demonstrated as suitable for use on very small sample quantities (c 10mg compared to 250mg for HPLC) which represents a distinct advantage over HPLC in many forensic scenarios, however as discussed in Chapter 4, there would be scope to reduce the sample quantities when using HPLC Peak Set A. Similarly, QGSTA is preferential to HPLC in cases where sample quantities are extremely low, since it requires only 50 grains of quartz per analysis.

The extensive body of research into the origin and persistence of quartz grain surface morphologies and the wax markers identified by GC in both in the forensic and geosciences literature has enabled the existence of soil databases containing wax marker and QGSTA data, which also enables these types of analysis to provide intelligence in "seek and find" cases. While this may be a perceived disadvantage of HPLC profiling compared to the existing techniques in some case scenarios, it does not diminish the applicability of the HPLC technique to the task of comparing and excluding close-proximity locations, for which it was intended.

There is significant potential for the combined use of both the HPLC and GC techniques for achieving very high accuracy rates for discriminating sample locations, and the accuracy achieved in this study using this combined approach was 100% across all sites and locations. This suggests that it would be beneficial to measure both types of markers where additional discrimination is required, providing there are sufficient sample quantities and resources available, and this approach would offer the additional benefit of the existing data on the wax-marker profiles of specific types of land use and vegetation. If HPLC were used to complement QGSTA, the combined benefits of each technique could allow for both the discrimination of close-proximity locations and geographical provenancing of samples, and since these are independent techniques, one set of results could be used to corroborate the other for maximum evidential value.

These comparative studies demonstrate that HPLC profiling has significant potential to provide investigators with a more accurate, cost-effective and simple alternative, to GC profiling of wax

markers and QGSTA, in scenarios where exclusionary analysis of close-proximity locations is desired, and would be a useful additional capability in laboratories where there established analyses are routinely performed.

8 Discussion

8.1 Introduction

This Chapter summarises the key outcomes of the research presented in this thesis and discuss the implications of the findings for forensic capabilities currently and identifies avenues for future research, where necessary. This thesis has presented findings within three overarching themes which recur throughout chapters 3-8. The first theme is the development of a practical HPLC sample preparation and analysis method for the purposes of discriminating samples from close proximity locations, which is necessary to ensure the final method could potentially be implemented in practice or have meaningful impact in the context of the almost entirely privatised market for forensic science service provision in England and Wales. The second theme relates to the suitability of the technique for general use in forensic casework, in different locations and at different points in time and times of year. The third theme pertains to the study of how the HPLC technique could be implemented in forensic casework with respect to regulatory requirements, for instance the use of validated and independent techniques. Summaries of the discussion on these themes is outlined in sections 8.1-8.4 and the key findings are highlighted in Table 8.1, while the areas identified for future research are discussed in section 8.5 and summarised in Table 8.2

Table 8.1 Summary of the Key Findings for each theme

Theme (Section)		Key Findings
Method Development (8.1)		Simplified sample preparation procedure reduces systematic error rate and cuts hard costs Reduced sample preparation time reduces costs and improves efficiency Data analysis strategy formalised for the first time. Use of CDFA provides objective method of discrimination Shown to have high accuracy rates at forensically relevant spatial scale
Generalisability	Spatial variability (8.2)	Highly accurate at discriminating close proximity locations Ability to discriminate is robust to changes in underlying geology No location markers were identified Ability to discriminate is robust to changes in season Ability to discriminate is robust to delays in sample analysis No temporal markers were identified
	Temporal Variability (8.3)	Ability to discriminate is robust to changes in season Ability to discriminate is robust to delays in sample analysis No temporal markers were identified
Implementation (8.4)		Choice of two alternative marker sets allows flexibility to suit case circumstances

	Superior discrimination at close-proximity than two existing techniques Compatible with an independent technique (QGSTA) allowing improved evidential value through combined analyses. Superior discrimination at close-proximity locations than two established techniques HPLC data improves the accuracy of discrimination by organic analysis when added to existing capabilities Adds value when used in conjunction with established techniques, allowing use of existing databases to narrow down search area in "seek and find" cases, with accurate close-proximity discrimination provided by HPLC.

8.1 Method Development

During this thesis, experiments were undertaken to determine the effects of various methods of handling, analysing and storing the samples (Chapter 3) (also discussed in McCulloch et al. (33)) The results obtained in this method development study indicate that it was not necessary to use the working concentration and sample amounts detailed in previous studies, therefore the sample concentration could be reduced to 500mg/ml and the amount required for analysis could be reduced to 250mg, which represents a fourfold reduction in comparison with previously cited sample quantities. Whilst it was not possible to determine the optimum storage conditions and packaging during this study through systematic experimental study, it was however ascertained that storage of samples in sealed containers exposed to light at ambient conditions was inappropriate due to the growth of mould, and the use of LDPE bags was not suitable for identification of sample components with LC-MS. On the basis of the results of the experiments testing the optimum sample preparation method, presented in sections 3.3.2.1, 3.3.2.4 and 3.3.2.5, it was not deemed necessary to use the sample preparation methods that theoretically could cause detrimental changes to the chromatographic profiles obtained, such as grinding, artificial drying or heating and the requirement for filtration could potentially be removed with further investigation as the cause of the pressure problems encountered.

It was therefore possible to simplify the sample preparation from the 10 steps reported by Bommarito et. al. (18) to just five steps, which reduces the opportunities for small human errors, such as inadvertent contamination of samples, inaccurate weighing or imprecise pipetting, to accumulate, therefore the number of opportunities for systematic error to occur in this newly developed method is also reduced by a factor of 50% relative to previous studies. As discussed in chapter 2, well designed analytical methods adopt a holistic approach to reducing quality risks, and the validation of forensic methods must consider and seek to

control the potential risk of error at each stage in the forensic process, therefore simple processes are desirable as they are easier, and therefore also quicker and cheaper, to validate for use in a commercial laboratory environment, in addition to being more efficient and potentially profitable for use in routine commercial analyses. In the current climate of forensic science provision in the UK, discussed in chapter 2, it is essential factors such as value for money and commercial viability are considered in forensic science research in order to achieve maximum impact in the wider forensic science community.

The newly developed method had greatly improved sensitivity compared to previous studies, and was approximately four times more sensitive than the method outlined by Bommarito et. al. (18) as it was possible to halve the amount of sample used, and then to halve the sample concentration without compromising the size of the peaks obtained or their signal to noise values. Previous forensic geoscience studies using HPLC required 5g (30) and 1g (31) (32) of soil, however this study proved it is possible to reduce the sample amount required to 250mg. It was also possible to reduce the analytical runtime from the 100 min, used by Bommarito et. al. (18), to 35mins by making use of gradient elution and new columns, and to the reduce the sample preparation time to less than one hour, compared to previous studies which took far longer at between approximately 3 hours (18) (31) (30) and approximately 16 hours (30). The new sample preparation method was also demonstrated to be more practical, with only five steps, than the established technique for the analysis of the organic fraction of soil in forensic geoscience, the determination of wax markers by GC (Chapter 7), for which the sample preparation involved in excess of 20 separate steps. In addition, the cost per sample using the new HPLC method was 90% less than the GC technique, furthermore the HPLC sample preparation was more efficient than the GC technique by a factor of 20.

Significant practical improvements have been outlined this thesis with regard to the processing and statistical analysis of the HPLC data generated with this new analytical methodology (Chapter 5), as presented by McCulloch et al (34). For the first time, specific sets of marker peaks have been identified that have been demonstrated to offer excellent discrimination over this close proximity, forensically relevant spatial scale, which allows for automatic integration of the chromatograms and, therefore, statistical analyses can be performed within minutes of analysis. Furthermore, two alternative marker sets were identified in this study, which could allow investigators to tailor any future analyses to case-specific requirements. The previous studies by Siegel and Precord (30) and Reuland and Trinler (31), relied on qualitative, visual comparison of the highly complex chromatograms obtained, which was naturally highly subjective and did not allow for quantitative, statistical assessment of differences between sample groups. Bommarito et. al. (18) used a combination of visual comparison, followed by

ad-hoc selection of peaks for further statistical analyses based on the results of the initial visual assessment. In the studies presented in this thesis it was also possible to visually identify peaks and peak height ratios that allowed the sample groups to be discriminated from one another for each of the data sets analysed (Chapters 4-7), however this was a highly subjective and time consuming approach, which could not be standardised or automated for general use with new data sets. None of the previous HPLC studies reported the length of time required to compare the chromatograms and assess the data, however the process of discriminating of samples by visual assessment of the chromatograms took several hours for the chromatograms generated in this study, therefore the ability to produce quantitative, statistical data in only a few minutes represents a significant improvement to the data analysis process. This new data analysis strategy was also demonstrated to be quicker and offer better discrimination than QGSTA (Chapter 7). The use of CDFA to quantitatively discriminate samples and provide a measure of statistical certainty to accompany the discriminant results reduces the subjectivity associated with the analytical results and, for the first time, it has been possible to verify that this method of HPLC profiling allows samples to be discriminated objectively with very high accuracy rates. The selection of these two sets of marker peaks not only creates a standardised approach that is much better suited for general use but also makes the data analysis for this new method far less labour intensive and a more reproducible process in comparison to the previous HPLC studies.

The method development and data analysis development described in Chapters 3 and 5 and presented in McCulloch et. al. (33) (34) therefore represent a substantial improvement to the overall practicality and potential impact of HPLC as a tool for forensic geoscientists relative to previously reported studies as it increases confidence with which discrimination of samples can be performed by allowing statistical analyses to be performed on the data very quickly and easily, it reduces the hard costs of the reagents and consumables required to analyse a sample, in addition to reducing the soft staffing costs related to preparing the samples for analysis and analysing the data. These improvements significantly increase the potential for this new HPLC technique to be successfully commercialised in the predominantly privatised forensic laboratories providing forensic services to the justice system in the UK.

8.2 Spatial Variability

The newly developed HPLC method presented in Chapter 4 and presented in McCulloch et. al. (34) was demonstrated to provide highly discriminatory results for samples obtained from close-proximity locations, which was considered more forensically relevant than the previous studies involving HPLC which only demonstrated the ability to discriminate locations that were

several kilometres apart (18) (30) (31) (32). The accuracy rate for the discriminant analysis obtained in the initial feasibility study (Chapter 4, McCulloch et al. (33)) was 100%, which compared very well to the results obtained in Chapter 7 using QGSTA, which was only able to discriminate one of the four locations at such close-proximity for this data set. Accuracy is a measure of the relative size of any discrepancy between the test result and a true result, and can only be calculated in circumstances where a true or accepted test result is available. The accuracy of a forensic test result can never be established with absolute certainty in casework, therefore care must be taken in court that the definition of the term accuracy is not misconstrued by the jury as infallibility, nonetheless in these controlled experiments the true location of origin was known therefore it is appropriate that accuracy rates have been calculated and reported.

The HPLC technique was also shown to perform better than the determination of wax markers by GC at this close-proximity spatial scale, providing higher rates of accuracy at this close-proximity scale for the majority of the data sets. The findings outlined in Chapter 6 confirm the potential for the general use of the new HPLC technique at different sites across the UK and internationally as the accuracy of discrimination was demonstrated to be high at multiple different locations. There were no consistent features observed in the profiles for samples from the same location type, nor for the samples from the same site, identified in chapter 6, therefore HPLC cannot yet be used in conjunction with land use and vegetation databases as there are currently no land use or geographical markers. The CDFA for both HPLC peak sets produced consistently high accuracy rates, at this close-proximity scale, at each the four sites discussed in Chapter 6 and by McCulloch et. al. (34), which demonstrates that the ability to discriminate is not limited to a specific geographical location and the technique is therefore robust to changes in the underlying geology of the sample site.

8.3 Temporal Variability

The new HPLC technique was demonstrated to provide very high discrimination of close-proximity sample locations at all of times of the year, for all the data sets studied (Chapter 6), which demonstrates that the ability of the HPLC technique to discriminate between close-proximity sample locations is not dependent upon the time of year in which the analysis is performed, and is therefore considered to be robust to seasonal changes. The prepared sample solutions were determined to be largely stable after two weeks, when stored at -20C (Chapter 3), however as discussed by McCulloch et. al. (33), significant changes were observed in the profiles after storage for 7 days in the fridge, and the physical samples were noted to be differentially affected under various storage conditions over time and it was determined that

LDPE bags and sealed glass containers caused mould-like artefacts in the samples after two weeks of storage (Section 3.3.3.2) therefore further investigation of shorter term variability of the profiles under different environmental conditions is required. The profiles were noted to vary at each specific sample location from one season to the next (Chapter 6), despite this, the ability to discriminate with extremely high rates of accuracy was maintained for peak set A, even in scenarios where samples were collected at different times of year from one another, which suggests that the HPLC new method is robust to operational delays in sample collection.

8.4 Implementation

In order to establish the suitability of this new HPLC method for implementation in casework, this thesis has not only outlined a practical and user friendly analytical method (Chapter 3), but also presented two alternative sets of marker peaks (Chapter 5) which are able to provide a flexible approach to data analysis which allows end users the choice of strategies to best suit the needs of a particular case. For example, the use of peak set A would be preferred if there have been delays between the time of the crime and the collection of reference samples, as this peak set is robust to such delays, however peak set B is far easier to compare visually, since there are fewer peaks to compare, therefore data analysis can be performed more quickly and with lower error rates, which would be advantageous should further, more advanced data analysis be required. With fewer marker peaks to investigate, characterise and optimise, the method could be quicker and easier to validate using peak set B as the target, meaning the technique could be used in casework sooner. The peaks in set A were deliberately selected from the peaks were consistently among the strongest contributors to the profiles at different sites and times of year (Chapter 5) and these peaks were able to be detected and quantified with good signal to noise values for all of the sample sets collected (Chapters 5-7), this means that the working concentration could be further reduced without losing the ability to reliably detect, identify and quantify the peaks of interest. Using peak set A, therefore, would be advantageous to investigators in cases where sample quantities are limited, as it could still be possible to obtain strong signals for peak set A, allowing for successful discrimination of close-proximity samples even in cases where less than 250mg of sample was available for comparison.

HPLC has been shown to be a useful addition to the techniques currently used in forensic geoscience, consistently offering high accuracy rates for the discrimination of close proximity sample locations, and was shown in Chapter 7 to offer better discrimination than both of the established techniques with which the results were compared, for the forensic scenario examined in this thesis. During the comparison with QGSTA, the HPLC allowed discrimination

of the sample from all four of the locations from one another, while QGSTA only allowed one location to be excluded from the others. When the performance of the new HPLC technique was compared with wax marker determination by GC, one of the few analytical techniques designed to study the organic fraction of soil that has been applied to forensic casework, the HPLC gave higher accuracy rates in most cases and the results could be generated much more quickly and cost effectively for the HPLC, offering significant potential for the HPLC to add value over the GC technique when implemented by a commercial or resource-limited service provider. It was not possible, however, to identify consistent features within the profiles that could be used as markers of a specific geographical location, type of vegetation or land use, nor were there any indicators of the season in which the samples were collected, which means that the HPLC cannot currently be used with databases in "seek and find" cases, and in this regard the QGSTA and GC approaches offer an advantage over the HPLC method presented in this thesis.

As discussed by McCulloch et. al. (33) (34) (35)and in Chapters 2 and 7, independent techniques are recommended for use in forensic casework, in order to strengthen the weight of evidence. Since HPLC has been shown to work well in conjunction with the independent analysis of the inorganic soil fraction by QGSTA. This would allow for the use of the two independent data sets obtained for HPLC and QGSTA to corroborate one another if implemented together in a forensic scenario. An added benefit of combining the use of HPLC and QGSTA is that this would permit the use of the existing sample databases in "seek and find" cases, with QGSTA allowing large geographical areas to be eliminated and HPLC providing the close-proximity precision to discriminate samples within the site of interest. Likewise, combined use of the GC and HPLC techniques would allow existing land use databases to be utilised with the GC data to narrow down search areas, in cases where it is necessary to determine provenance in the absence of a known location of interest against which to make sample comparisons, with HPLC providing the discrimination of any close-proximity sample locations. Implementing the HPLC and GC techniques together could simultaneously allow the use of databases and improve the overall accuracy with which the organic fraction can be used to discriminate at this close-proximity spatial scale since, in the few cases where HPLC was unable to discriminate all of the samples, the addition of the wax marker data set to the CDFA improved the accuracy to 100%, and vice versa. The speed and cost-effectiveness with which this new HPLC technique provides highly accurate discriminant analyses also suggests that it shows excellent potential for use as a screening tool and could, for instance, be used to eliminate those samples that can be discriminated by HPLC, reducing the number of samples required to be analysed with the more expensive and time consuming techniques, that are

more acceptable for presentation in court, and thereby improving the overall efficiency of the analysis.

8.5 Further Work

The suggestions highlighted in this thesis for potentially useful future studies, to build on the findings produced during the work for this thesis, are summarised in Table 8.2

Table 8.2 Summary of Suggestions for Future Research

Theme (Section)		Key Findings
Method Development (8.1)		Perform stability studies to optimise the packaging and storage conditions Investigate the effects of grinding on intra-location variability Investigate the effects of further reducing the working concentration and sample amount.
Generalisability	Spatial variability (8.2)	Identify land use markers through analysis of the existing data set Collect samples from a wider range of land use types to identify potential markers for additional land use types Collect samples using more strictly controlled sample sites to enable easier identification of potential land use markers
	Temporal Variability (8.3)	Identify the markers most stable over time through analysis of the existing data set Identify markers most suitable for discriminating samples collected at the same site at different times through analysis of the existing data set Collect samples more frequently to identify temporal trends Collect samples more frequently to establish the stability of the existing markers over shorter time periods Collect samples over extended time periods to establish the long term marker stability. Collect additional samples to identify markers that best discriminate samples collected at specific times of year
Implementation (8.4)		Isolate, purify and characterise marker compounds for use as reference standards Identify marker compounds through spectroscopic analyses Validate the accuracy, precision, robustness and linear range of the HPLC method for the identification and quantification of peak ,marker sets A and B. Investigate matrix effects by extracting HPLC profiles from a range of evidence types Determine the stability of the HPLC profiles on various types of evidence under different environmental conditions Conduct transfer and persistence studies to determine the ability to recover and detect peaks of interest in forensically relevant scenarios.

Further studies are required to optimise and validate this new HPLC technique for use in forensic case work. As discussed in section 8.1, it is necessary to optimise the method of storage and packaging for the physical samples, to ascertain the stability of the samples when refrigerated or frozen in different types of packaging, and to verify that the paper containers used in this study do not introduce artefacts as the LDPE bags and glass vials were observed to. It may also be appropriate to grind the individual replicate samples taken at each of the locations, and it is therefore recommended that the effect of homogenising the samples is investigated to ascertain whether this would improve the intra-location variability, and consequently the accuracy and confidence levels obtained in the statistical analysis. Careful consideration of the effects of grinding the sample on the ability to correctly interpret the data obtained from the sample would be necessary prior to introducing this step into the sample preparation procedure, since homogenisation causes the irrevocable loss of information on the internal variability of the composition of any given soil sample, which is necessary for the correct interpretation of data obtained by other techniques, such as QGSTA, therefore there may be practical and operational implications of grinding the samples prior to analysis, such as changes to the order in which independent or complementary analyses are conducted. Peak set A consistently produced relatively large peaks in this study, therefore there is potential to use this method with smaller sample quantities, and so it is also recommended that further experiments are conducted into the effect of reducing the working concentration on the signal to noise values of the peaks of interest.

As this is the first time this HPLC method has been used, it is not possible to identify marker compounds for specific types of vegetation through comparison of the chromatography obtined in this study with previous studies published in the literature. It may be possible, however, with further experimental research, to identify individual or groups of peaks which are indicative of specific land uses or vegetation types (section 8.2), by collecting additional samples from a wider range of land uses, but with more strictly defined criteria for the vegetation and land use of the sample groups used in future studies, for instance by selecting sites with a limited range of types of vegetation eg. agricultural or forestry commission land. Likewise, it may also be possible to select seasonal markers or identify temporal trends through further studies utilising a more frequent sampling strategy (section 8.3), which would aid interpretation in cases where there has been a delay between time of the crime and the collection of evidence and reference samples. Alternatively, it may be possible to utilise the temporal variability observed for the two marker sets in this study to demonstrate the ability of the HPLC technique to discriminate samples collected at different time points in future

studies, or by performing additional analyses on the existing data set. In addition to generating new data through further sample collection and analysis, it could be possible to identify marker peaks better suited for examining temporal or land uses differences between comparator samples by generating an algorithm to re-examine the complete data set generated for this thesis, with the specific aim of selecting the chromatographic features most stable over time, or indicative of specific times of year, or land use types. It is therefore recommended that further research should be conducted to identify land use markers for use in conjunction with existing soil databases to enable the new HPLC method to be used in "seek and find" intelligence cases.

In order to ultimately implement the HPLC technique in forensic casework in the UK, it will be necessary to validate the technique in accordance with the regulator's guidelines, which would involve extensive investigation of both source and activity level propositions. It is first necessary to establish that the HPLC method is suitably accurate and precise to identify and quantify the marker peaks, by preparing isolating and purifying extracts of the marker compounds and performing method validation on the current HPLC method, to establish the linear range of the method and ensure reliable measurement of the marker peaks in future analyses. Further characterisation of the peaks of interest is also required, for instance through spectroscopic analysis of purified extracts, to enable confirmation of their identity in future studies, for instance through the comparison of the UV spectra collected by the HPLC using the current method, or by incorporating a subsequent MS detection step. In addition, it will be necessary to determine any relevant matrix effects by examining the profiles obtained for samples recovered from various evidence types eg. clothing and footwear, and the stability of the profiles under different environmental conditions over time must also be established in future studies. Furthermore, empirical studies should be conducted into the transfer and persistence of soil, and any effects on the resulting HPLC profiles, on various surfaces and in case-specific conditions, in order to aid interpretation at the activity level in forensically relevant scenarios.

9 Conclusions

This thesis had seven aims, which were presented in Chapter 2 and here in table 9.1

Table 9.1 Summary of objectives and conclusions.

Objective	Conclusion	See Section
1. To determine whether it is possible to discriminate soil samples obtained from different locations within the same site	It was possible to discriminate close-proximity soil samples, as high rates of accuracy were achieved using the new analytical methodology	Feasibility Study (Chapter 4)
2. To reduce the quantity of soil required to prepare the HPLC sample	A fourfold reduction in soil quantity, relative to previous studies, was achieved	Method Development (Chapter 3)
3. To simplify the sample preparation procedure	The sample preparation was simplified from 10 steps to 5 steps, relative to previous studies	Method Development (Chapter 3)
4. To reduce the time required to complete each analysis	The preparation time was reduced to <1 hour, compared to 3-16 hours in previous studies Data Analysis now possible in minutes and semi-automated	Method Development (Chapter 3) Data analysis development (Chapter 5)
5. To apply the technique to samples from locations representing a range of underlying geologies	Discrimination by HPLC was highly accurate at a sites of different underlying geologies	Geographic and Seasonal Variability (Chapter 6)
6. To monitor temporal variations in soil chromatography	Discrimination by HPLC was highly accurate at different times of year Discrimination by HPLC was robust to delays in sampling	Geographic and Seasonal Variability (Chapter 6)
7. To compare the discriminatory power of the method with existing geoforensic techniques	HPLC offered higher average rates of accuracy for discriminating close-proximity sites than wax marker determination by GC HPLC offered more accurate discrimination than Quartz Grain Surface Texture Analysis at close proximity locations	Complementary Analyses (Chapter 7) Complementary Analyses (Chapter 7)

9.1 Aim 1: To determine whether it is possible to discriminate soil samples obtained from different locations within the same site

This first aim was essential to augmenting current forensic geoscience capabilities and providing investigators with a means of discriminating closely situated sample sites and improving upon the spatial resolution offered by previous studies. This research presented in this thesis demonstrated that 100% accuracy in discriminating soil from different close-proximity locations was achieved in the feasibility study (Chapter 4), and high levels of accuracy were also achieved throughout the discriminant analyses performed in Chapters 5-7. Therefore this thesis shows it is indeed possible to use HPLC profiling to distinguish soils from different locations within the same site, which has not been demonstrated in any of the previous studies in the published literature.

9.2 Aim 2: To reduce the quantity of soil required to prepare the HPLC sample

This aim was critical to achieve in order that the technique was more applicable to trace evidence analysis and relevant to a wider range of forensic scenarios, where only small amounts of sample are available. The findings presented in Chapter 3 addressed this aim demonstrating that it was possible to reduce the sample amount to 250mg, compared to the smallest amount quoted in the literature, which was 1g.

9.3 Aim 3: To simplify the sample preparation procedure

The third aim was to simplify the sample preparation in order to reduce the minimise opportunities for human or systematic errors, such as the contamination of samples, and this aim was also achieved during method development (Chapter 3) as the newly developed method had only five steps compared to the most recent method in the literature, which had 10 steps.

9.4 Aim 4: To reduce the time required to complete each analysis

The fourth aim of the project was to reduce the amount of time required to prepare and analyse the HPLC samples, in order to better meet the needs of the predominantly commercial providers of forensic science services in England and Wales, and this was achieved during method development (Chapter 3) and the development of data analysis strategies (Chapter 5). The sample preparation time was reduced from between 3 and 16 hours in previous studies to less than one hour, and the analysis time was reduced to 35min compared to 100min in

previous studies in Chapter 3. The previous papers did not specify how long the data analysis step took, however all utilised visual comparison as a preliminary step to discriminating samples, which is highly subjective and labour intensive, and in Chapter 4, semi-automated data analysis was possible through the use of pre-selected sets of marker peaks, therefore data analysis could be performed in a single step in SPSS, which resulted in a reduction of the time required for data analysis from two hours for visual comparison of the data obtained for the feasibility study (Chapter 4) to approximately five minutes. The overall analysis time was therefore significantly reduced relative to the previous studies and the newly developed method represents a significant improvement to the status quo.

9.5 Aim 5: To apply the technique to samples from locations representing a range of underlying geologies

The fifth aim of the project was to apply the technique to samples from locations representing a range of underlying geologies, in order to establish whether the ability to discriminate was restricted to a particular geographical location or type of underlying geology and assess the ability to apply the technique across different geographical areas. This aim was achieved during the development of data analysis strategies (Chapter 5) and the examination of geographical variability (Chapter 6). It was demonstrated that the ability of the newly developed HPLC technique to achieve highly accurate discrimination of close proximity samples was not affected by the underlying geology of a site and could indeed be applied at a range of different sites within the UK and internationally, therefore the technique was robust to differences in underlying geology.

9.6 Aim 6: To monitor temporal variations in soil chromatography

The sixth aim addressed in this research was to apply the technique at different times of year to further assess the generalisability of the newly developed HPLC method, and this aim was achieved during the examination of (Chapter 6), where the technique was demonstrated to achieve high levels of accuracy at all of the time points tested in this study. The newly developed technique was not only able to discriminate close proximity locations accurately throughout the year, but was also robust to delays in sample collection, which is beneficial to the implementation of the technique in forensic casework where there can be no control over the length of time between a crime being committed and being detected.

9.7 Aim 7: To compare the discriminatory power of the method with existing geoforensic techniques

The final aim of this thesis was to compare the ability of the newly developed techniques to existing techniques, in order to establish the potential for the newly developed technique to add value to forensic geoscience cases. The aim to compare the ability of the newly developed technique to offer additional benefits over established techniques at discriminating samples from close proximity locations was met in Chapter 7, which investigated complementary uses of the technique. The new HPLC technique was found to offer better discrimination of close proximity locations than the quartz grain surface texture analysis, which is independent of the organic analysis by HPLC, and could only exclude one of the four locations examined in this study, while HPLC was able to discriminate all four locations. In addition HPLC was shown to offer benefits over the determination of wax markers by GC, which is an established technique currently used for the analysis of the organic fraction of soil in UK courts, offering higher average accuracy rates at the majority of sites examined in this study. This shows that HPLC can add value relative to both established independent techniques, adding to the weight of evidence that forensic soils and sediments can provide to investigators.

9.8 Summary

The research outlined in this thesis has achieved the primary aim of developing an HPLC method for the discrimination of close proximity, trace soil and sediment samples for forensic use, which has been achieved through the seven objectives detailed above. It has been possible to reduce the sample quantity required by 75%, and to double the sensitivity of the analysis. The efficieny of the HPLC analysis was improved by 186% which, along with the simplified sample preparation, reduced the cost of performing the analysis by 90%. This new HPLC technique offers small-scale resolution that was previously unavailable, allowing investigators the ability to discriminate soils with better spatial precision than the analytical techniques previously presented in forensic geoscience literature, and was also shown to offer a higher rate of accuracy than two established techniques when discriminating soils from8u544 the close-proximity locations used in this study.

The findings of this research are important for two key reasons, firstly the demonstration of the excellent ability of HPLC to allow discrimination at close proximity locations provided in this thesis greatly improves potential impact of the technique through the possibility of applying forensic geoscience in a wider range of cases, and secondly the consideration of the needs of end-users during the method development and experimental design stages show that

it is possible to undertake meaningful forensic science research that is also attractive to commercial forensic service providers in the context of the continued reduction in police spending on forensic science.

For the first time it has been demonstrated that HPLC analysis allows for highly accurate discrimination of trace quantities of soil in the forensically relevant scenario in which the locations of interest are situated within the same site or spatially constrained area, where the underlying geology is unlikely to vary and therefore traditional techniques based on mineralogy and elemental analysis may be less useful. The technique was shown to offer benefits when used in conjunction with existing, established techniques, enhancing the quality and utility of the intelligence offered by forensic soil and sediment analysis through allowing the advantages of the extremely high levels of accuracy and small scale spatial resolution offered by the HPLC method to be combined with the use of the soil and land use databases created using more established techniques to assist provenance determination in "seek and find" cases .

This work has therefore expanded the range of forensic scenarios to which forensic geoscience techniques can be applied, and added to the variety of crime reconstruction challenges that the information afforded from the analysis of soils and sediments can assist investigators in tackling. Furthermore, through ensuring that the operational needs of practitioners and the commercial concerns of other relevant stakeholders have been addressed and prioritised appropriately, in light of the serious financial and regulatory constraints faced by forensic science practitioners in the UK, this thesis presents sound scientific research that not only offers a novel approach to forensic geoscience and crime reconstruction but also maximises the potential impact of the research findings by developing a truly practical, user-friendly and commercialisable analytical technique suitable for use by a wider variety of forensic science stakeholder organisations

10 References

1. Ruffell A, McKinley J. Geoforensics Chichester: John Wiley & Sons Ltd; 2008.

2. Stam M. Soil as Significant Evidence in a Sexual Assault/Attempted Homicide Case. In Pye K, Croft DJ, editors. Forensic Geoscience: Principles, Techniques and Applications. London: Geological Society; 2004. p. 295-299.

3. Palenik. Forensic Miocroscopy. Microscopy and Analysis. 1997 November: p. 7-9.

4. Ruffell A, McKinley J. Geoforensics Chichester: John Wiley & Sons Ltd; 2008

5. Brown AG, Smith A, Elmhurst O. The Combined use of Pollen and Soil Analyses in a Search and Subsequent Murder Investigation. Journal of Forensic Sciences. 2002; 47(3): p. 614-618.

6. Lombardi G. The Contribution of Forensic Geology and Other Trace Evidence Analysis to the Investigation of the Killing of Italian Prime Minister Aldo Moro. Journal of Forensic Sciences. 1999; 44(3): p. 634-642.

7. Murray RC. Forensic Geology: Yesterday, Today and Tomorrow. In Pye K, Croft DJ, editors. Forensic Geoscience: Principles, Techniques and Applications. London: Geological Society; 2004. p. 7-9.

8. Murray RC. Evidence from the Earth Forensic Geology and Criminal Investigation Missoula: Mountain Press; 2004.

9. Anonymous. Forensic Examination of Soil Evidence. In 13th INTERPOL Forensic Science Symposium; 2001; Lyon, France. p. D1 175-D1 191.

10. Sugita R, Suzuki S, Y. K. Forensic Geology- A Review: 2004-2007. In 15th INTERPOL Forensic Science Symposium; 2007; Lyon, Framce. p. 81-85.

11. Sugita R, Yoshida H. Forensic Geology: Review 2007 to 2009. In 16th International Forensic Science Symposium Interpol; 2010; Lyon, France.

12. Morgan RM., Bull PA. The Use of Grain Size Distribution Analysis of Sediments and Soils in Forensic Enquiry. Science and Justice. 2007; 47(3): p. 125-135.

13. Morgan RM., Bull PA. Forensic Geoscience and Crime Detection. Identification, Interpretation and Presentation in Forensic Geoscience. Minerva Medicolegale. 2007; 127(2): p. 73-89

14. Morgan RM, Bull PA. The Philosophy, Nature and Practice of Forensic Sediment Analysis. Progress in Physical Geography. 2007; 31(1): p. 43-58.

15. Bull PA, Morgan RM, Freudiger-Bonzon J. A Critique of the Present Use of Some Geochemical Techniques in Geoforensic Analysis. Forensic Science International. ; 178(2-3): p. e35-e40.

16. Morgan R, Cohen J, McGookin I, Murly-Gotto J, O'Connor R, Muress S, Freudiger-Bonzon J, Bull PA. The relevance of the evolution of experimental studies for the interpretation and evaluation of some trace physical evidence. Science and Justice. 2009; 49(4): p. 277-285.

17. Bergslien E. An introduction to forensic geoscience Chichester: Wiley-Blackwell; 2012.

18. Bommarito CR, Sturdevant AB, Symanski DW. Analysis of Forensic Soil Samples Via High-Performance Liquid Chromatography and Ion Chromatography. Journal of Forensic Sciences. 2007; 52(1): p. 24-30.

19. Soil-net. www.soil-net.com. [Online].; 2016. Available from: http://www.soil-net.com/dev/page.cfm?pageid=secondary_intro_recognition&loginas=anon_secondary

20. Fitzpatrick RW. Forensic Comparison of Soils. In Tibbett M, Carter DO, editors. Soil Analysis in Forensic Taphonomy: Chemical and Biological Effects of Buried Remains. Boca Raton: CRC Press; 2008. p. 1-28.

21. Minasny B, McBratney AB, Salvador-Blanes S. Quantitative Models for Pedogenesis- A Review. Geoderma. 2008; 144(1-2): p. 140-157.

22. Gallop A, Stockdale R. Trace and Contact Evidence. In White P, editor. Crime Scene to Court the Essentials of Forensic Science. Cambridge: Royal Society of Chemistry; 1998. p. 47-72.

23. Ahmad UK, Rajendran S. Forensic Analysis of Explosive Residues from Hand Swabs. Malaysian Journal of Analytical Sciences. 2008; 12(1): p. 25-31.

24. Tarvin M, McCord B, Mount K, Miller ML. Analysis of hydrogen peroxide field samples by HPLC/FD and HPLC/ED. Forensic Science International. 2011; 209(1-3): p. 166-172.

25. Ide R. Fire Investigation. In White P, editor. Crime Scene to Court the Essentials of Forensic Science. Cambridge: Royal Society of Chemistry; 1998. p. 133-158.

26. Hodgson B. Alcohol: Blood. In Encyclopedia of Forensic Sciences. Cambridge: Academic Press; 2000. p. 74-80.

27. Pragst F, Balikova MA. State of the Art in Hair Analysis for Detection of Drug and Alcohol Abuse. Clinica Chimica Acta. 2006; 370(1-2): p. 17-49.

28. Pizzolato TM, Lopez de Alda MJ, Barcelo D. LC-Based Analysis of Drugs of Abuse and their Metabolites in Urine. TrAC Trends in Analytical Chemistry. 2007; 26(6): p. 609-624.

29. Gaillard Y, Breuil R, Doche C, Romeuf L, Lemeur C, Prevosto JM, et al. Detection of Amitriptyline, Nortriptyline and Bromazepam in Liver, CSF and Hair in the Homicidal Poisoning of a One Month Old Girl Autopsied Eight Months After Death. Forensic Science International. 2011; 207(1-3): p. e16-e18.

30. Siegel JA, Precord C. The Analysis of Soil Samples by Reverse Phase High Performance Liquid Chromatography Using Wavelength Ratioing. Journal of Forensic Sciences. 1985; 30(2): p. 511-525.

31. Reuland DJ, Trinler WA. An Investigation of the Potential of High Performance Liquid Chromatography for the Comparison of Soil Samples. Forensic Science International. 1981; 18: p. 201-208.

32. Reuland DJ, Trinler WA, Farmer MD. Comparison of Soil Samples by High Performance Liquid Chromatography Augmented by Absorbance Ratioing. Forensic Science International. 1992; 52(2): p. 131-142.

33. McCulloch G, Morgan RM, Bull PA. High Performance Liquid Chromatography as a valuable tool for geoforensic soil analysis. Australian Journal of Forensic Sciences. (DOI 10.1080/00450618.2016.1194474) Published online 2016 June 22.

34. McCulloch G, Brewer M, Dawson LA, Morgan RM. The identification of markers for Geoforensic HPLC profiling at close proximity sites. Forensic Science International. 2017; 272: p127-141.

35. McCulloch G, Dawson LA, Ross JM, Morgan RM. The discrimination of geoforensic trace material from close proximity locations by organic profiling using HPLC and plant wax marker analysis by GC. Forensic Science International (In Submission).

36. Kirk PL. The ontogeny of criminalistics. Journal of Criminal Law, Criminology and Police Science. 1963;(54): p. 235-238.

37. Locard E. L'enquête criminelle et les méthodes scientifiques Paris: Flammarion; 1920

38. DeForest P, Gaensslen R, Lee H. Forensic Science: An Introduction to Criminalistics New York: McGraw Hill; 1983.

39. Osetburg JW. The Crime Laboratory: Case Studies of Scientific Criminal Investigation : Indiana University Press; 1968.

40. Kirk PL. Crime Investigation. 2nd ed. New York: Wiley; 1974.

41. Inman K, Rudin N. The origin of evidence. Forensic Science International. 2002 January;(126): p. 11-16.

42. Murray R, Tedrow JCF. Forensic Geology Engelwood Cliffs: Prentice Hall; 1992.

43. Chisum JW, Turvey BE. Evidence Dynamics: Locard's Exchange Principle and Crime Reconstruction. Journal of Behavioural Profiling. 2000 January; 1(1).

44. French JS, Morgan RM, Bull PA, Baxendel P. Multiple Transfers of Particulates and their Dissemination Within Contact Networks. Science and Justice. 2012; 52(1): p. 33-41.

45. King S, Benson S, Kelly T, Lennard C. Determining the effects of routine fingermark detection techniques on the subsequent recovery and analysis of explosive residues on various substrates. Forensic Science International. 2013; 323(1-3): p. 257-264.

46. Abdul-Karim N, Morgan R, Binions R, Harrison K, Temple T. The spatial distribution of post-blast RDX residue: forensic implications. Journal of Forensic Sciences. 2013; 58(2): p. 365-371.

47. Edelman GJ, Roos M, Bolck A, Aalders MC. Practical Implementation of Blood Stain Age Estimation Using Spectroscopy. IEEE Journal of Selected Topics in quantum electronics. 2016; 22(3)

48. Association of Chief Police Officers. _connect Forensic Science KTN. [Online].; 2012 [cited 2016 April 12. Available from: https://connect.innovateuk.org/web/forensics/article-view/-/blogs/harnessing-science-and-innovation-for-forensic-investigation-in-policing?_33_redirect=https%3A%2F%2Fconnect.innovateuk.org%2Fweb%2Fforensics%2Farticle-view%3Fp_p_id%3D101_INSTANCE_TLbDGoIbFrk0%26p_p_lifecycle%3D0%26p_p_state%3Dnormal%26p_p_mode%3Dview%26p_p_col_id%3Dcolumn-2%26p_p_col_pos%3D1%26p_p_col_count%3D2

49. O'Sullivan S, Geddes T, & Lovelock TJ. The migration of fragments of glass from the pockets to the surfaces of clothing. Forensic Science International. 2011; 208(1-3): p. 149-155.

50. Morgan RM, French JC, O'Donnell L, Bull PA. The reincorporation and redistribution of trace geoforensic particulates on clothing: An introductory study. Science & Justice. 2010; 50(4): p. 195-199.

51. Finnebraaten M, Granér T, Hoff-Olsen P. May a speaking individual contaminate the routine DNA laboratory? Forensic Science International:Genetics Supplement Series. 2008; 1(1): p. 421-422.

52. Poy A, van Oorschot RAH. Beware; gloves and equipment used during the examination of exhibits are potential vectors for transfer of DNA-containing material. International Congress Series. 2006; 1288: p. 556-558.

53. Committee on Identifying the Needs of the Forensic Sciences Community, National Research Council. Strengthening Forensic Science in the United States: A Path Forward. Washington DC: The National Academies of Sciences, Engineering and Medicine, National Research Council; 2009.

54. Silverman B. Research and Development in Forensic Science: a Review. Home Office, Forensic Science Regulator; 2011.

55. Silverman B. Review of Research and Development in Forensic Science: University Responses. Home Office, Forensic Science Regulator; 2011.

56. Silvermann B. Review of Research and Development in Forensic Science: Other responses. Home Office, Forensic Science Regulator; 2011.

57. Sangshettia JN, Deshpandea M, Zaheera Z, Shindeb DB, Arotec R. Quality by design approach: Regulatory need. Arabian Journal of Chemistry (DOI 10.1016/j.arabjc.2014.01.025). Published online 2014 February 10.

58. International Conference of Harmonisation. Q8 Pharmaceutical Development. ICH, Quality Guidelines; 2009.

59. International Organisation for Standardisation. ISO/IEC 17025:2005(en) General requirements for the competence of testing and calibration laboratories. ISO/IEC, Committee on conformity assessment; 2005.

60. Forensic Science Regulator. www.gov.uk. [Online]; 2014 [cited 16 June 02]. Available from: https://www.gov.uk/government/uploads/attachment_data/file/375285/FSR-G-201_Validation_guidance_November_2014.pdf.

61. International Laboratory Accreditation Cooperation. Modules in a Forensic Science Process IL AC-G19:08/2014. [Online].; 2014 [cited 2016 June 02]. Available from: http://ilac.org/latest_ilac_news/ilac-g19082014-published/.

62. McKinley J. How useful are databases in environmental and criminal forensics? Geological Society Special Publication. 2013; 384: p. 109-119.

63. Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA. A model for case assessment and interpretation. Science and Justice. 1998; 38: p. 151–156.

64. Cook R, Evett IW, Jackson G, Jones PJ, Lambert JA. A hierarchy of propositions: deciding which level to address in casework. Science and Justice. 1998; 38: p. 231–239.

65. Jackson G, Jones S, Booth G, Champod C, Evett IW. The nature of forensic science opinion – a possible framework to guide thinking and practice in investigations and in court proceedings. Science and Justice. 2006; 46: p. 33-44.

66. Daubert v. Merrell Dow Pharmaceuticals. 509 U.S.579. 1993 June 28;(92-102).

67. Law Commission. Expert Evidence in Criminal Proceedings in England and Wales. Law Commission. 2011; LC325.

68. Walport M, Craig C, Surkovic E. Annual Report of the Government Chief Scientific Adviser 2015: Forensic Science and Beyond: Authenticity, Provenance and Assurance. Evidence and Case Studies. Government Office for Science; 2015.

69. The Committee on the Judiciary House of Representatives. Rule 702 Testimony by Expert Witness. Federal Rules of Evidence. 2015 December 1; Article VII: p. 15.

70. House of Commons Science and Technology Committee. Forensic Science on Trial Seventh Report of Session 2004–05 Volume II Oral and written evidence. ; 2005.

71. House of Commons Science and Technology Committee. Forensic Science on Trial Seventh Report of Session 2004–05 Report, together with formal minutes. ; 2005.

72. House of Commons Science and Technology Committee. The Forensic Science Service Seventh Report of Session 2010–12 Volume I: Report, together with formal minutes, oral and written evidence. ; 2011.

73. House of Commons Science and Technology Committee. The Forensic Science Service Seventh Report of Session 2010–12 Volume II Additional written evidence. ; 2011.

74. McFarland R. Review of the Forensic Science Service. Home Office; 2003.

75. Home Office. Obtaining better value from police procurement: The Police Act 1996 (equipment) Regulations 2010 And The Police Act 1996 (services) Regulations 2010. Home Office, Police Productivity Unit; 2010.

76. Penning M. www.parliament.uk. [Online].; 2016 [cited 2016 April 18]. Available from: https://www.parliament.uk/business/publications/written-questions-answers-statements/written-statement/Commons/2016-03-11/HCWS612/.

77. Cranfield Soil and Agrifood Institute. Cranfield University. [Online].; 2016 [cited 07 June 2016]. Available from: https://www.cranfield.ac.uk/case-studies/research-case-studies/national-soil-map.

78. The James Hutton Institute. The James Hutton Institute. [Online].; 2016 [cited 2016 August 04]. Available from: http://sifss.hutton.ac.uk

79. Spector C. Soil Science Education, Biospheric Sciences Research. [Online].; 2012 [cited 2012 June 01]. Available from: http://soil.gsfc.nasa.gov/index.php?section=88.

80. Hillel D. Introduction to Environmental Soil Physics London: Academic Press; 2004.

81. Jenny H. Factors of Soil Formation A system of Quantitative Pedology. New York: Dover Publications Inc.; 1994.

82. Fitzpatrick EA. Soils : their formation, classification and distribution. London: Longman; 1980.

83. Tan KH. Principles of soil chemistry. Boca Raton: CRC Press; 1998.

84. O'Hare G, Rooke A. Soils, vegetation, ecosystems. Edinburgh: Oliver and Boyd; 1988.

85. Kolay AK. Soil Genesis, Classification and Evaluation Volume 1. New Delhi: Atlantic; 2007.

86. Esu IE. Soil Characterisation, Classification and Survey. Ibadan: HEBN Publishers Plc.; 2010.

87. Ashman MR, Puri G. Essential Soil Science: a clear and concise introduction to soil science. Oxford: Blackwell Science; 2002.

88. Nichols G. Sedimentology and Stratigraphy. Oxford: Blackwell Science Ltd.; 1999.

89. Moore RCp3. Meaning of Facies. Sedimentary facies in geologic history. The Geological Society of America Memoir Baltimore: Waverley Press Inc; 1949.

90. Lewis DG, McConchie D. Practical Sedimentology. London: Chapman and Hall; 1994.

91. Pye K. Geological and Soil Evidence: Forensic Applications. Boca Raton: CRC Press; 2007.

92. Dawson LA, Hillier SS. Measurement of Soil Characteristics for Forensic Applications. Surface and Interface Analysis. 2010; 42: p. 363-377.

93. Hazen RM, Ferry JM. Mineral Evolution: Mineralogy in the Fourth Dimension. Elements. 2010: 6 (1) p. 9-12.

94. Sverjensky DA, Lee N. The Great Oxidation Event and Mineral Diversification. Elements. 2010: 6 (1) p. 31-36.

95. Johnson DL, Keller EA, Rockwell TK. Dynamic Pedogenesis: New Views on Some Key Soil Concepts, and a model for Interpreting Quaternary Soils. Quaternary Research. 1990; 33(3): p. 306-319.

96. Miami College of Arts and Sciences. Department of Biology. [Online]. [cited 14 07 2016]. Available from: http://www.bio.miami.edu/dana/160/160S16_22print.html.

97. Saferstein R. Physical properties: Glass and Soil. In Criminalistics: An Introduction to Forensic Science. Upper Saddle River: Prentice Hall; 2000.

98. Morgan RM, Gibson A, Little M, Hicks L, Dunkerley S, Bull PA. The Preservation of quartz grain surface textures following vehicle fire and their use in forensic enquiry. Science and Justice. 2008; 48(3): p. 133-140.

99. Sugita R, Marumo Y. Unique Particles in Soil Evidence. In Pye K, Croft DJ, editors. Forensic Geoscience: Principles, Techniques and Applications. London: Geological Society; 2004. p. 97-102.

100. Nederbragt AJ, Dunbar RB, Osborn AT, Palmer A, Thurow J, Wagner T. Digital Sediment Colour Analysis from Digital Images and Correlation with Sediment Composition. In Francus P, editor. Image Analysis Sediments and Paleoenvironments. Dordrecht: Kluwer Academic Publishers; 2004. p. 105-124.

101. Petraco N, Kubic T. A Density Gradient Technique for Use in Forensic Soil Analysis. Journal of Forensic Sciences. 2000; 45(4): p. 872-873.

102. Junger EP. Assessing the Unique Characteristics of Close-Proximity Soil: Just How Useful is Soil? Journal of Forensic Sciences. 1996; 41(1): p. 27-34.

103. Sugita R, Marumo y. Screening of Soil Evidence by a Combination of Simple Techniques: Validity of Particle Size Distribution. Forensic Science International. 2001; 122(2-3): p. 155-158.

104. Blott SJ, Croft DJ, Pye K, Saye SE, Wilson HE. Particle Size Analysis by Laser Diffraction. In Pye K, Croft DJ, editors. Forensic Geoscience: Principles, Techniques and Applications. London: Geological Society; 2004. p. 63-73.

105. Pye K, Blott SJ. Particle size analysis of sediments, soils and related particulate materials for forensic purposes using laser granulometry. Forensic Science International. 2004; 144(1): p. 19-27.

106. Morgan RM, Wiltshire P, Parker A, Bull PA. The Role of Forensic Geoscience inWildlife Crime Detection. Forensic Science International. 2006; 16(1-3): p. 152-162.

107. Morgan RM, Bull PA. Data interpretation in forensic sediment and soil geochemistry. Environmental Forensics. 2006; 7(4): p. 325-333.

108. Dawson LA, Campbell CD, Hillier S, Brewer MJ. Methods of Characterising and Fingerprinting Soils for Forensic Application. In Tibbett M, Carter DO, editors. Soil Analysis in Forensic Taphonomy. Chemical and Biological Effects of Buried Human Remains. Boca Raton: CRC Press; 2008. p. 271-307.

109. Dawson LA, Towers W, Mayes RW, Craig J,V. The use of plant hydrocarbon signatures in characterising soil organic matter. In Pye K, Croft DJ. Forensic Geoscience: Principles, Techniques and Applications. London: The Geological Society of London; 2004. p. 269-276.

110. Dawson LA, Mayes RW. Criminal and environmental soil forensics: Soil as physical evidence in forensic investigations. In Murphy BL, Morrison RD, editors. Introduction to environmental forensics. 3rd ed. Oxford: Academic Press; 2014. p. 457-486.

111. Mayes RW, Macdonald LM, Ross JM, Dawson LA. Discrimination of Domestic Garden Soils Using Plant Wax Compounds as Markers. In: Ritz K, Dawson LA, Miller D (Ed.) Criminal and Environmental Soil Forensics. Berlin/Heidelberg: Springer; 2009. p. 463-474.

112. Morrison AR, McColl SM, Dawson LA, Brewer MJ. Characterisation and Discrimintaion of Urban Soils: Preliminary Results from the Soil Forensics University Network. In: Ritz K, Dawson LA, Miller D (Ed.) Criminal and Environmental Soil Forensics. Berlin/Heidelberg: Springer; 2009. p. 75-86.

113. Croft DJ, Pye K. Stable Carbon and Nitrogen Isotope Variations in Soils: Forensic Applications. In Croft DJ, Pye K, editors. Forensic Geoscinece: Techniques, Principles and Applications.: Geological Society; 2004. p. 257-267.

114. Wright RJ. Atomic Absorption and Flame Emission Spectrometry. In Sparks DL, Page AL, Helmke PA, Loeppert RH, Soltanpour PN, Tabatabai MA, editors. Methods of Soil Analysis Part 3. Chemical Methods. Madison: Soil Science Society of America Inc. and American Society of Agronomy Inc.; 1996. p. 65-90.

115. Jarvis KE, Wilson HE, James SL. Assessing Element Variability in Small Soil Samples Taken During Forensic Investigations. In Pye K, Croft DJ, editors. Forensic Geoscience: Principles, Techniques and Applications. London: Geological Society; 2004. p. 171-182.

116. Pye K, Blott SJ, Wray DS. Elemental Analysis of Soil Samples for Forensic Purposes byInductively Coupled Plasma Spectrometry – Precision Considerations. Forensic Science International. 2006; 160: p. 178-192.

117. Cheshire K, Morgan RM, Holmes J. The potential for geochemical discrimination of single and mixed source soil samples from close proximity urban parkland locations. Australian Journal of Forensic Sciences (DOI 10.1080/00450618.2016.1144789). Published online 2016 April 05.

118. Espana JS, Ercilla MD. Geochemical Modelling of Concentrated Mine Waters: A Comparison of the Pitzer Ion-Interaction Theory with the Ion-Association Model for the Study of Melanterite Solubility in San Telmo Mine: In Stefansson O, editor. Geochemistry Research Advances. New York: Nova Science Publishers Inc; 2008. p. 31-55.

119. Russell RJ. Infrared Methods. In Wilson MJ, editor. A Handbook of Determinative Methods in Clay Mineralogy. Glasgow and London: Blackie; 1987. p. 133-173.

120. Pye K. Forensic Examination of Rocks, Sediments and Soils and Dusts Using Scanning Electron Microscopy and X-ray Chemical Microanalysis. In Pye K, Croft DJ, editors. Forensic Geoscience: Principles, Techniques and Applications. London: Geological Society; 2004. p. 103-121.

121. Young JM, Weyrich LS, Cooper A. Forensic soil DNA analysis using high-throughput sequencing: A comparison of four molecular markers. FSI Genetics. 2014; 13: p. 176-184.

122. Meyers MS, Foran RD. Spatial and Temporal Influences on Bacterial Profiling of Forensic Soil Samples. Forensic Science International. 2008; 53(3): p. 652-660.

123. Mildenhall DC, Wiltshire PE, Bryant VM. Forensic Palynology: Why do it and how it works. Forensic Science International. 2006 November 22; 163(3): p. 163-172.

124. Morgan RM, Allen E, King T, Bull PA. The spatial and temporal distribution of pollen in a room: Forensic implications. Science and Justice. 2014; 54(1): p. 49-56.

125. Scott RK, Morgan RM, Jones V, Cameron NG. The transferability of diatoms to clothing and the methods appropriate for their collection and analysis in forensic geoscience. Forensic Science International. 2014 August; 241: p. 127-137.

126. Walls HJ. Forensic Science: An Introduction to Scientific Crime Detection. 2nd ed. London: Sweet and Maxwell; 1974.

127. Houck M. Mute Witnesses: Trace Evidence Analysis New York CIty: Academic Press Inc; 2001.

128. Houck M. Trace Evidence Analysis: More Cases in Mute Witnesses: Elsevier Academic Press; 2004.

129. Isphording WC. The Right Way and the Wrong Way of Presenting Statistical and Geological Evidence in a Court of Law (a little knowledge is a dangerous thing!). In Pye K, Croft DJ, editors. Forensic Geoscience: Principles, Techniques and Applications. London: Geological Society; 2004. p. 281-288.

130. Bull PA, Morgan RM. Sediment Fingerprints: A Forensic Technique Using Quartz Sand Grains…A response. Sciece and Justice. 2007; 47(3): p. 141-144.

131. Bull PA, Morgan RM, Sagovsky A, Hughes GJA. The Transfer and Persistence of Trace Particulates: Experimental Studies Using Clothing Fabrics. Science and Justice. 2006; 46(3): p. 185-195.

132. Bull P, Morgan R. Book Review: Geological and Soil Evidence, Forensic Applications, K.Pye, 2007, CRC Press, London. Science and Justice. 2008; 48(4): p. 200-202.

133. Pye K. Sediment fingerprints: a Forensic Technique Using Quartz Sand Grains--a Comment. Science and Justice. 2007; 47(1): p. 34-36.

134. Cohen LE, Felson M. Social Change and Crime Rate Trends. American Sociological Review. 1979; 44(August): p. 588-608.

135. Bull PA, Parker A, Morgan RM. The Forensic Analysis of Soils and Sediment Taken from the Cast of a Footprint. Forensic Science International. 2006; 162(1-3): p. 6-12.

136. Equitas. Innovative Contributions to the Documentation of Missing Persons and Clandestine Cemetaries in Colombia's Conflict. ; 2006.

137. Bull PA, Morgan RM. Sediment Fingerprints: A forensic technique using quartz sand grains. Science and Justice. 2006; 46(2): p. 107-124.

138. Torrent J, Schwertmann U, Fechter H, Alferez F. Quantitative Relationships Between Soi Colour and Hematite Content. Soil Science. 1983; 136(6): p. 354-358.

139. Rawlins BG, Cave M. Investigating Multi-Element Soil Geochemical Signatures and their Potential for use in Forensic Studies. In Pye K, Croft DJ, editors. Forensic Geoscience: Principles, Techniques and Applications. London: Geological Society; 2004. p. 197-206.

140. Pounds C, Smalldon K. The Transfer of Fibres Between Clothing Materials During Simulated Contacts and their Persistence During Wear: Part I — Fibre Transference. Journal of the Forensic Science Society. 1975; 15: p. 17-27.

141. Pounds C, Smalldon K. The Transfer of Fibres Between Clothing Materials During Simulated Contacts and their Persistence During Wear: Part II — Fibre Transference. Journal of the Forensic Science Society. 1975; 15: p. 29-37.

142. Pounds C, Smalldon K. The Transfer of Fibres Between Clothing Materials During Simulated Contacts and their PersistenceDuring Wear: Part III —A Preliminary Investigation of the Mechanisms Involved. Journal of the Forensic Science Society. 1975; 15: p. 197-207.

143. Peldszus. Organic Acids Analysis in Environmental Samples, Ion Chromatography for Determination of. In Encyclopedia of Analytical Chemistry.: John Wiley and Sons; 2006.

144. Ahuja S, Dong MW, editors. Handbook of Pharmaceutical Analysis by HPLC Oxford: Elsevier Academic Press; 2005.

145. Waters. How Does High Performance Liquid Chromatography Work. [Online].; 2012 [cited 2012 July 20]. Available from: http://www.waters.com/waters/nav.htm?cid=10049080&locale=en_GB.

146. Cole M. Drugs of Abuse. In White P. Crime Scene to Court, The Essentials of Forensic Science. Cambridge: Royal Society of Chemistry; 1998. p. 210-231.

147. Atkins P, de Paula J. Molecular Symmetry. In Atkins P, de Paula J. Physical Chemistry. 9th ed. Oxford: Oxford university Press; 2010.

148. Hansen CM. Hansen solubility parameters: a user's handbook Boca Raton: CRC Press; 2007.

149. Snyder LR, Dolan J, Jupille TH. A condensed course in the systematic development of liquid chromatography separtions. In HPLC Method Development; 2011; Edinburgh: Hichrom Ltd.

150. Snyder LR, Kirkland JJ, Dolan J. Introduction to Modern Liquid Chromatography Hoboken: John Wiley and Sons; 2010.

151. Snyder LR, Kirkland JJ, Glajch JL. Practical HPLC Method Development: John Wiley and Sons; 2012.

152. International Council for Harmonisation of Technical Requirements for Pharmaceuticals for Human Use. www.ich.org. [Online].; 2005 [cited 2016 July 18]. Available from: http://www.ich.org/products/guidelines/quality/article/quality-guidelines.html.

153. Ambach L, Penitschka F, Broillet A, König S, Weinmann W, Bernhard W. Simultaneous quantification of delta-9-THC, THC-acid A, CBN and CBD in seized drugs using HPLC-DAD. Forensic Science International. 2014 October; 243: p. 107-111.

154. Müller IB, Windberg CN. Validation of an HPLC Method for Quantitation of MDMA in Tablets. Journal of Chromatographic Science. 2005; 43(8): p. 434-437.

155. Sultan M. Simultaneous HPLC determination and validation of amphetamine, methamphetamine, caffeine, paracetamol and theophylline in illicit siezed tablets. Intrnational Journal of Pharmacy and Pharmaceutical Sciences. 2014; 6(4): p. 294-298.

156. United Nations Office on Drugs and Crime, Laboratory Scientific Section. unodc.org. [Online].; 2009 [cited 2016 August 12]. Available from: https://www.unodc.org/documents/scientific/ST-NAR-40-Ebook.pdf.

157. Floriania G, Gasparetto JC, Pontarolo R, Gonçalves AG. Development and validation of an HPLC-DAD method for simultaneous determination of cocaine, benzoic acid, benzoylecgonine and the main adulterants found in products based on cocaine. Forensic Science International. 2014; 235: p. 32-39.

158. Lurie IS, Bailey CG, Anex DS, Bethea MJ, McKibben TD, Casale JF. Profiling of impurities in illicit methamphetamine by high performance liquid chromatography and capillary electrochromatography. Journal of Chromatography A. 2000;(870): p. 53-68.

159. United nations office on drugs and crime, Laboratory and Scientific Section. unodc.org. [Online].; 2005 [cited 2016 August 12]. Available from: https://www.unodc.org/pdf/publications/report_st-nar-35.pdf.

160. Fernández P, Morales L, Vázquez C, Bermejo AM, Tabernero MJ. HPLC–DAD determination of opioids, cocaine and their metabolites in plasma. Forensic Science International. 2006 August 10; 161(1).

161. Kristoffersen L, Strand DH, Liane VH, Vindenes V, Tvete IF, Aldrin M. Determination of safety margins for whole blood concentrations of alcohol and nineteen drugs in driving under the influence cases. Forensic Science International. 2016 February ; 259: p. 119-126.

162. Sporkert F, Brunel C, Augsburger MP, Mangin P. Fatal tolperisone poisoning: Autopsy and toxicology findings in three suicide cases. Forensic Science International. 2012 February 10; 215(1-3): p. 101-104.

163. Yan H, Zhou X, Shen B, Xiang P, Shen M. Determination of Nitrite in whole blood by High Performance Liquid Chromatography. Journal of Forensic Sciences. 2015 August 14; 61(1): p. 254-258.

164. Xu Y, Zhou X, Shi X. HPLC and HPLC/MS analysis of red ink on counterfeit 100-yuan notes. Forensic Science International. 2016 February; 259: p. 47-52.

165. Morgan SL. Validation of Forensic Characterization and Chemical Identification of Dyes Extracted from Millimeter-length Fibers. US Department of Justice, National Institute of Justice; 2014. Report No.: DUNS 11-131-0249.

166. Afiq M, Huri M, Ahmad UK. Forensic analysis of high explosive from selected cloth. The Malaysian Journal of Analytical Sciences. 2014; 18(1): p. 68-77.

167. Abdul-Karim N, Blackmann CS, Gill PP, Karu K. The spatial distribution patterns of condensed phase post-blast explosive residues formed during detonation. Journal of Hazardous Materials. 2016 October 5; 316: p. 204-213.

168. Simmons K. Soil Sampling SESDPROC-300-R3. [Online].; 2014 [cited 2016 June 07]. Available from: https://www.epa.gov/quality/quality-system-and-technical-procedures-sesd-field-branches.

169. Larson SA. Digital Commons. [Online].; 2012 [cited 2012 June 19]. Available from: http://digitalcommons.unl.edu/cgi/viewcontent.cgi?article=1008&context=biochemdiss.

170. Newell AJ, Morgan RM, Griffin LD, Bull PA, Marshall JR, Graham G. Automated Texture Recognition of Quartz Sand Grains for Forensic Applications. Journal of Forensic Sciences. 2012 September; 57(5): p. 1285-1289.

171. Dove H, Mayes RW. Protocol for the analysis of n-alkanes and other plant wax compounds and for their use as markers for quantifying the nutrient supply of large mammalian herbivores. Nature Protocols. 2006; p. 1680-1696.